

# An Achievable Rate Region for the Broadcast Channel with Feedback

Ramji Venkataramanan

Yale University

ramji.venkataramanan@yale.edu

S. Sandeep Pradhan \*

University of Michigan, Ann Arbor

pradhanv@eecs.umich.edu

January 20, 2013

## Abstract

A single-letter achievable rate region is proposed for the two-receiver discrete memoryless broadcast channel with generalized feedback. The coding strategy involves block-Markov superposition coding using Marton's coding scheme for the broadcast channel without feedback as the starting point. If the message rates in the Marton scheme are too high to be decoded at the end of a block, each receiver is left with a list of messages compatible with its output. Resolution information is sent in the following block to enable each receiver to resolve its list. The key observation is that the resolution information of the first receiver is correlated with that of the second. This correlated information is efficiently transmitted via joint source-channel coding, using ideas similar to the Han-Costa coding scheme. Using the result, we obtain an achievable rate region for the stochastically degraded AWGN broadcast channel with noisy feedback from only one receiver. It is shown that this region is strictly larger than the no-feedback capacity region.

## 1 Introduction

The two-receiver discrete memoryless broadcast channel (BC) is shown in Figure 1(a). The channel has one transmitter which generates a channel input  $X$ , and two receivers which receive  $Y$  and  $Z$ , respectively. The channel is characterized by a conditional law  $P_{YZ|X}$ . The transmitter wishes to communicate information simultaneously to the receivers at rates  $(R_0, R_1, R_2)$ , where  $R_0$  is the rate of the common message, and  $R_1, R_2$  are the rates of the private messages of the two receivers. This channel has been studied extensively. The largest known set of achievable rates for this channel without feedback is due to Marton [1]. Marton's rate region is equal to the capacity region in all cases where it is known. (See [2], for example, for a list of such channels.)

Figure 1(b) shows a BC with generalized feedback.  $S_n$  represents the feedback signal available at the transmitter at time  $n$ . This model includes noiseless feedback from both receivers ( $S_n = (Y_n, Z_n)$ ), partial feedback ( $S_n = Y_n$ ) as well as noisy feedback ( $S_n = Y_n + \text{noise}$ ). El Gamal showed in [3] that feedback does not enlarge the capacity region of a physically degraded BC. Later, through a simple example, Dueck [4] demonstrated that feedback can strictly improve the capacity region of a general BC. For the stochastically

---

\*This work was supported by NSF grants CCF-0915619 and CCF-1111061. It was presented in part at the IEEE International Symposium on Information Theory (ISIT) 2010, held in Austin, TX.

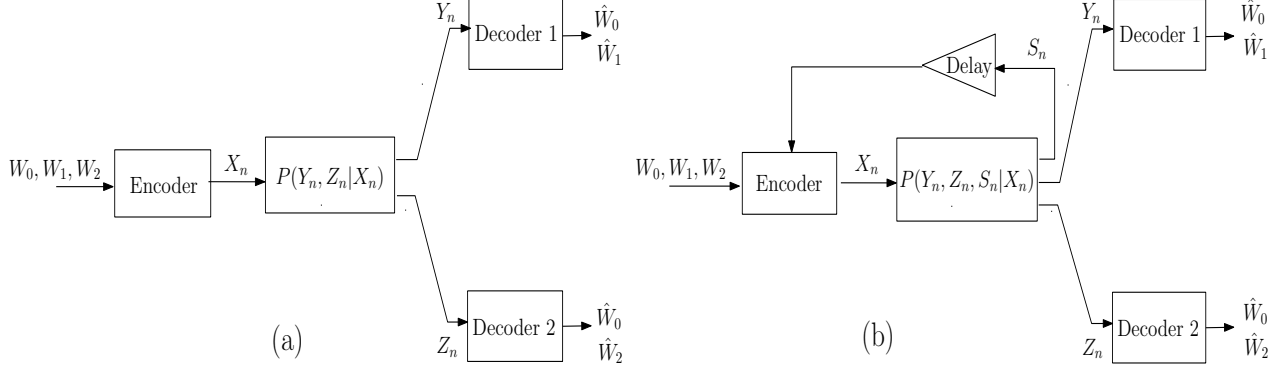


Figure 1: The discrete memoryless broadcast channel with a) no feedback b) generalized feedback.

degraded AWGN broadcast channel with noiseless feedback, an achievable rate region larger than the no-feedback capacity region was established in [5], and more recently, in [6]. A finite-letter achievable rate region (in terms of directed information) for the discrete memoryless BC with feedback was obtained by Kramer [7]; using this characterization, it was shown that rates strictly outside the no-feedback capacity region could be achieved for the binary symmetric BC with noiseless feedback.

In this paper, we establish a single-letter achievable rate region for the memoryless BC with generalized feedback. We use the proposed region to compute achievable rates for the stochastically degraded AWGN BC with noisy feedback from one receiver, and show that rates strictly outside the no-feedback capacity region can be achieved.

Before describing our coding strategy, let us revisit the example from [4]. Consider the BC in Figure 2. The channel input is a binary triple  $(X_0, X_1, X_2)$ .  $X_0$  is transmitted cleanly to both receivers. In addition, receiver 1 receives  $X_1 \oplus N$  and receiver 2 receives  $X_2 \oplus N$ , where  $N$  is an independent binary Bernoulli( $\frac{1}{2}$ ) noise variable. Here, the operation  $\oplus$  denotes the modulo-two sum. Without feedback, the maximum sum rate for this channel is 1 bit/channel use, achieved by using the clean input  $X_0$  alone. In other words, no information can be reliably transmitted through inputs  $X_1$  and  $X_2$ .

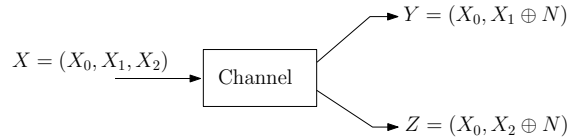


Figure 2: The channel input is a binary triple  $(X_0, X_1, X_2)$ .  $N \sim \text{Bernoulli}(\frac{1}{2})$  is an independent noise variable.

Dueck described a simple scheme to achieve a greater sum rate using feedback. In the first channel use, transmit one bit to each receiver  $i$  through  $X_i$ ,  $i = 1, 2$ . Receivers 1 and 2 then receive  $Y = X_1 \oplus N$  and  $Z = X_2 \oplus N$ , respectively, and cannot recover  $X_i$ . The transmitter learns  $Y, Z$  through feedback and can compute  $N = Y \oplus X_1 = Z \oplus X_2$ . For the next channel use, the transmitter sets  $X_0 = N$ . Since  $X_0$  is received noiselessly by both receivers, receiver 1 can now recover  $X_1$  as  $Y \oplus N$ . Similarly, receiver 2 reconstructs  $X_2$  as  $Z \oplus N$ . We can repeat this idea over several transmissions: in each channel use, transmit a fresh pair of bits (through  $X_1, X_2$ ) as well as the noise realization of the previous channel use (through  $X_0$ ). This yields a sum rate of 2 bits/channel use. This is, in fact, the sum-capacity of the channel since it equals the cut-set bound

$\max_{P_X} I(X; YZ)$ .

The example suggests a natural way to exploit feedback in a broadcast channel. If we transmit a block of information at rates outside the no-feedback capacity region, the receivers cannot uniquely decode their messages at the end of the block. Each receiver now has a list of codewords that are jointly typical with its channel output. In the next block, we attempt to resolve these lists at the two receivers. The key observation is that the resolution information needed by receiver 1 is in general *correlated* with the resolution information needed by receiver 2. The above example is an extreme case of this: the resolution information of the two receivers is identical, i.e., the correlation is perfect!

In general, the two receivers' resolution information are not perfectly correlated, but can still be transmitted over the BC more efficiently than independent information. This is analogous to transmitting correlated sources over a BC using joint source-channel coding [8–12]. At the heart of the proposed coding scheme is a way to represent the resolution information of the two receivers as a pair of correlated sources, which is then transmitted efficiently in the next block using joint source-channel coding, along the lines of [8]. We repeat this idea over several blocks of transmission, with each block containing independent fresh information superimposed over correlated resolution information for the previous block.

The following are the main contributions of this paper:

- We obtain a single-letter achievable rate region for the discrete memoryless BC with generalized feedback. The proposed region contains three extra random variables in addition to those in Marton's rate region.
- Using a simpler form of the rate region with only one extra random variable, we compute achievable rates for the AWGN broadcast channel with noisy feedback. It is shown that rates outside the no-feedback capacity region can be achieved even with noisy feedback from only one receiver. This is the first characterization of achievable rates for the AWGN BC with noisy feedback at finite SNR, and is in contrast to the finding in [13] that noisy feedback does not increase the prelog of the sum-capacity as the SNR grows asymptotically large.

One feature of the proposed region is that it includes the case where there a common message to be transmitted to both receivers, in addition to their private messages. The previously known schemes for the AWGN BC with noiseless feedback [5, 6] assume that there is no common message.

- At the conference where our result was first presented [14], another rate region for the BC with feedback was proposed independently by Shayevitz and Wigger [15]. Though a direct comparison of the two regions does not appear feasible, we show that the rates for the examples presented in [15] can also be obtained using the proposed region.

*Notation:* We use uppercase letters to denote random variables, lower-case for their realizations and calligraphic notation for their alphabets. Bold-face notation is used for random vectors. Unless otherwise stated, all vectors have length  $n$ . Thus  $\mathbf{A} \triangleq A^n \triangleq (A_1, \dots, A_n)$  represents a random vector, and  $\mathbf{a} \triangleq a^n \triangleq (a_1, \dots, a_n)$  a realization. The  $\epsilon$ -strongly typical set of block-length  $n$  of a random variable with distribution  $P$  is denoted  $\mathcal{A}_\epsilon^{(n)}(P)$ .  $\delta(\epsilon)$  is used to denote a generic positive function of  $\epsilon$  that goes to zero as  $\epsilon \rightarrow 0$ . Logarithms are with base 2, and entropy and mutual information are measured in bits. For  $\alpha \in (0, 1)$ ,  $\bar{\alpha} \triangleq 1 - \alpha$ .  $\oplus$  denotes modulo-two addition.

In the following, we give an intuitive description of a two-phase coding scheme for communicating over a BC with noiseless feedback. We will use the notation  $\sim$  to indicate the random variables used in the first

phase. Thus  $(\tilde{Y}, \tilde{Z})$  denotes the channel output pair for the first phase, and  $(Y, Z)$  the output pair for the second phase. We start with Marton's coding strategy for the discrete memoryless BC without feedback. The message rates of the two receivers are assumed to lie outside Marton's achievable rate region. Let  $\tilde{U}$ ,  $\tilde{V}$ , and  $\tilde{W}$  denote the auxiliary random variables used to encode the information.  $\tilde{W}$  carries the information meant to be decoded at both receivers.  $\tilde{U}$  and  $\tilde{V}$  carry the rest of the information meant for the receivers 1 and 2, respectively. The  $\tilde{U}$ - and  $\tilde{V}$ -codebooks are constructed by randomly sampling the  $\tilde{U}$ - and  $\tilde{V}$ -typical sets, respectively. Let  $\tilde{\mathbf{U}}$ ,  $\tilde{\mathbf{V}}$  and  $\tilde{\mathbf{W}}$  denote the three random codewords chosen by the transmitter. The channel input vector  $\tilde{\mathbf{X}}$  is obtained by 'fusing' the triple  $(\tilde{\mathbf{U}}, \tilde{\mathbf{V}}, \tilde{\mathbf{W}})$ .

Since the rates lie outside Marton's region, the receivers are not able to decode the information contained in  $\tilde{U}$ ,  $\tilde{V}$ , and  $\tilde{W}$ . Instead, they can only produce a list of highly likely codewords given their respective channel output vectors. At the first decoder, this list is formed by collecting all  $(\tilde{U}, \tilde{W})$ -codeword pairs that are jointly typical with the channel output. A similar list of  $(\tilde{V}, \tilde{W})$ -codeword pairs is formed at the second receiver. Note that even with feedback, the total transmission rate of the BC cannot exceed the capacity of the point-to-point channel with input  $\tilde{X}$  and outputs  $(\tilde{Y}, \tilde{Z})$  (since the channel is memoryless). Hence, given *both* channel output vectors  $(\tilde{\mathbf{Y}}, \tilde{\mathbf{Z}})$ , the posterior probability of the codewords will be concentrated on the transmitted codeword triple.

At the end of the first phase, the feedback vector  $\tilde{\mathbf{S}}$  is available at the encoder. In the second phase, we treat  $(\tilde{\mathbf{U}}, \tilde{\mathbf{W}})$  as the source of information to be transmitted to the first decoder, and  $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}})$  as the source of information to be transmitted to the second decoder. The objective in the second phase is to communicate these two correlated pairs over the BC, while treating  $\tilde{\mathbf{S}}$  as source state information and  $\tilde{\mathbf{Y}}$  and  $\tilde{\mathbf{Z}}$  as side-information available at the two receivers. This is accomplished using a joint source-channel coding strategy. Transmission of correlated information over a BC has been addressed in [8, 11].

In the Han-Costa framework [8], the correlated information is modeled as a pair of memoryless sources characterized by a fixed single-letter distribution. The pair of sources is first covered using codebooks constructed from auxiliary random variables; the covering codewords are then transmitted over the BC using Marton coding. The current setup differs from [8] in two ways. First, the correlated information given by  $(\tilde{\mathbf{U}}, \tilde{\mathbf{W}})$  and  $(\tilde{\mathbf{V}}, \tilde{\mathbf{W}})$  does not exhibit a memoryless-source-like behavior. This is because the vectors  $\tilde{\mathbf{U}}$ ,  $\tilde{\mathbf{V}}$  and  $\tilde{\mathbf{W}}$  come from codebooks. However, when the codewords are sufficiently long and are chosen randomly,  $(\tilde{\mathbf{U}}, \tilde{\mathbf{V}}, \tilde{\mathbf{W}})$  will be jointly typical and can be covered using auxiliary codebooks similar to [8]. The second difference from [8] is the presence of source state information  $\tilde{\mathbf{S}}$  and side-information  $\tilde{\mathbf{Y}}$  and  $\tilde{\mathbf{Z}}$  available at receivers 1 and 2, respectively. We handle this by extending both the covering and channel coding steps of the Han-Costa scheme to incorporate the side-information. Thus at the end of the second phase, the decoders are able to decode their respective messages.

We will superimpose the two phases using a block-Markov strategy. The overall transmission scheme has several blocks, with fresh information entering in each block being decoded in the subsequent block. The fresh information gets encoded in the first phase, and is superimposed on the second phase which corresponds to information that entered in the previous block.

It turns out that the performance of such a scheme cannot be directly captured by single-letter information quantities. This is because the state information, given by the channel outputs of all the previous blocks, keeps accumulating, leading to a different joint distribution of the random variables in each block. We address this issue by constraining the distributions used in the second phase (Definition 2.3) so that in every block, all the sequences follow a stationary joint distribution. This results in a first-order stationary Markov process of the

sequences across blocks.

The rest of paper is organized as follows. In Section 2, we define the problem formally and state the main result of the paper, an achievable rate region for BC with generalized feedback. We give an outline of the proof of the coding theorem in Section 3. In Section 4, we use the proposed region to compute achievable rates for the AWGN BC with noisy feedback. We also compare our region with the one proposed by Shayevitz and Wigger. The formal proof of the coding theorem is given in Section 5, and Section 6 concludes the paper.

## 2 Problem Statement and Main Result

A two-user discrete memoryless broadcast channel with generalized feedback is a quintuple  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{Y\mathcal{Z}\mathcal{S}|X})$  of input alphabet  $\mathcal{X}$ , two output alphabets  $\mathcal{Y}, \mathcal{Z}$ , feedback alphabet  $\mathcal{S}$  and a set of probability distributions  $P_{Y\mathcal{Z}\mathcal{S}|X}(\cdot|x)$  on  $\mathcal{Y} \times \mathcal{Z} \times \mathcal{S}$  for every  $x \in \mathcal{X}$ . The channel satisfies the following conditions for all  $n = 1, 2, \dots$

$$\Pr(Y_n = y_n, Z_n = z_n, S = s_n | X^n = \mathbf{x}, Y^{n-1} = \mathbf{y}, Z^{n-1} = \mathbf{z}, S^{n-1} = \mathbf{s}) = P_{Y\mathcal{Z}\mathcal{S}|X}(y_n, z_n, s_n | x_n) \quad (1)$$

for all  $(y_n, z_n, s_n) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{S}$ ,  $\mathbf{x} \in \mathcal{X}^n$ , and  $(\mathbf{y}, \mathbf{z}, \mathbf{s}) \in \mathcal{Y}^{n-1} \times \mathcal{Z}^{n-1} \times \mathcal{S}^{n-1}$ . The schematic is shown in Figure 1(b). We note that the broadcast channel with noiseless feedback from both receivers is a special case with  $\mathcal{S} = \mathcal{Y} \times \mathcal{Z}$ , and  $S_n = (Y_n, Z_n)$ .

**Definition 2.1.** An  $(n, M_0, M_1, M_2)$  transmission system for a given broadcast channel with generalized feedback consists of

- A sequence of mappings for the encoder:

$$e_m : \{1, 2, \dots, M_0\} \times \{1, 2, \dots, M_1\} \times \{1, 2, \dots, M_2\} \times \mathcal{S}^{m-1} \rightarrow \mathcal{X}, \quad m = 1, 2, \dots, n, \quad (2)$$

- A pair of decoder mappings:

$$g_1 : \mathcal{Y}^n \rightarrow \{1, 2, \dots, M_0\} \times \{1, 2, \dots, M_1\}, \quad g_2 : \mathcal{Z}^n \rightarrow \{1, 2, \dots, M_0\} \times \{1, 2, \dots, M_2\}. \quad (3)$$

**Remark:** Though we have defined the transmission system above for feedback delay 1, all the results in this paper hold for feedback with any finite delay  $k$ .

We use  $W_0$  to denote the common message, and  $W_1, W_2$  to denote the private messages of decoders 1 and 2, respectively. The messages  $(W_0, W_1, W_2)$  are uniformly distributed over the set  $\{1, 2, \dots, M_0\} \times \{1, 2, \dots, M_1\} \times \{1, 2, \dots, M_2\}$ . The channel input at time  $n$  is given by  $X_n = e_n(W_0, W_1, W_2, S^{n-1})$ . The average error probability of the above transmission system is given by

$$\tau = \frac{1}{M_0 M_1 M_2} \sum_{k=1}^{M_0} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \Pr((g_1(Y^n), g_2(Z^n)) \neq ((k, i), (k, j)) \mid (W_0, W_1, W_2) = (k, i, j)). \quad (4)$$

**Definition 2.2.** A triple of non-negative real numbers  $(R_0, R_1, R_2)$  is said to be achievable for a given broadcast channel with feedback if  $\forall \epsilon > 0$ , there exists an  $N(\epsilon) > 0$  such that for all  $n > N(\epsilon)$ , there exists an  $(n, M_0, M_1, M_2)$  transmission system satisfying the following constraints:

$$\frac{1}{n} \log M_0 \geq R_0 - \epsilon, \quad \frac{1}{n} \log M_1 \geq R_1 - \epsilon, \quad \frac{1}{n} \log M_2 \geq R_2 - \epsilon, \quad \tau \leq \epsilon. \quad (5)$$

The closure of the set of all achievable rate pairs is the capacity region of the channel.

We now define the structure for the joint distribution of all the variables in our coding scheme. Due to the block-Markov nature of the scheme, the random variables carrying the resolution information in each block depend on the variables corresponding to the previous block. In order to obtain a single-letter rate region, we need the random variables in each block to follow the same joint distribution, say  $P$ . Hence, after each block of transmission, we generate the variables for the next block using a Markov kernel  $Q$  that has invariant distribution  $P$ . This will guarantee a stationary joint distribution  $P$  in each block.

**Definition 2.3.** Given a broadcast channel with feedback  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{YZS|X})$ , define  $\mathcal{P}$  as the set of all distributions  $P$  on  $\mathcal{U} \times \mathcal{V} \times \mathcal{A} \times \mathcal{B} \times \mathcal{C} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{S}$  of the form

$$P_{ABC} P_{UV|ABC} P_{X|ABCUV} P_{YZS|X},$$

where  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{U}$ , and  $\mathcal{V}$  are arbitrary sets. Consider two sets of random variables  $(U, V, A, B, C, X, Y, Z, S)$  and  $(\tilde{U}, \tilde{V}, \tilde{A}, \tilde{B}, \tilde{C}, \tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{S})$  each having the same distribution  $P$ . For brevity, we often refer to the collection  $(A, B, S)$  as  $K$ , to  $(\tilde{A}, \tilde{B}, \tilde{S})$  as  $\tilde{K}$ , and to  $\mathcal{A} \times \mathcal{B} \times \mathcal{S}$  as  $\mathcal{K}$ . Hence

$$P_{\tilde{U}\tilde{V}\tilde{K}\tilde{X}\tilde{Y}\tilde{Z}} = P_{UVCKXYZ} = P.$$

For a given  $P \in \mathcal{P}$ , define  $\mathcal{Q}(P)$  as the set of conditional distributions  $Q$  that satisfy the following consistency condition

$$P_{ABC}(a, b, c) = \sum_{\tilde{u}, \tilde{v}, \tilde{k}, \tilde{c} \in \mathcal{U} \times \mathcal{V} \times \mathcal{K} \times \mathcal{C}} Q_{ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C}}(a, b, c | \tilde{u}, \tilde{v}, \tilde{k}, \tilde{c}) P_{UVKC}(\tilde{u}, \tilde{v}, \tilde{k}, \tilde{c}), \quad \forall (a, b, c). \quad (6)$$

Then for any  $P \in \mathcal{P}$  and  $Q \in \mathcal{Q}(P)$ , the joint distribution of the two sets  $(U, V, K, C, X, Y, Z)$  and  $(\tilde{U}, \tilde{V}, \tilde{K}, \tilde{C}, \tilde{X}, \tilde{Y}, \tilde{Z})$  is

$$P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}\tilde{X}\tilde{Y}\tilde{Z}} Q_{ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C}} P_{UVKXYZ|ABC}. \quad (7)$$

With the above definitions, we have the following theorem.

**Theorem 1.** Given a broadcast channel with generalized feedback  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{YZS|X})$ , for any distribution

$P \in \mathcal{P}$  and  $Q \in \mathcal{Q}(P)$ , the convex hull of the following region is achievable.

$$R_0 < \min\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4, \mathcal{T}_5\} \quad (8)$$

$$R_0 + R_1 < I(\tilde{U}AC; Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K}; AC|\tilde{U}\tilde{C}) \quad (9)$$

$$R_0 + R_2 < I(\tilde{V}BC; Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K}; BC|\tilde{V}\tilde{C}) \quad (10)$$

$$R_0 + R_1 + R_2 < I(\tilde{U}AC; Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K}; AC|\tilde{U}\tilde{C}) - \mathcal{T} \quad (11)$$

$$+ I(\tilde{V}; C|\tilde{C}) + I(\tilde{V}B; Z\tilde{Z}\tilde{B}|C\tilde{C}) - I(\tilde{U}\tilde{K}A; B|C\tilde{V}\tilde{C})$$

$$R_0 + R_1 + R_2 < I(\tilde{V}BC; Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K}; BC|\tilde{V}\tilde{C}) - \mathcal{T} \quad (12)$$

$$+ I(\tilde{U}; C|\tilde{C}) + I(\tilde{U}A; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{C}) - I(\tilde{V}\tilde{K}B; A|C\tilde{U}\tilde{C})$$

$$2R_0 + R_1 + R_2 < I(\tilde{U}AC; Y\tilde{Y}\tilde{A}|\tilde{C}) - I(\tilde{V}\tilde{K}; AC|\tilde{U}\tilde{C}) - \mathcal{T} \quad (13)$$

$$+ I(\tilde{V}BC; Z\tilde{Z}\tilde{B}|\tilde{C}) - I(\tilde{U}\tilde{K}; BC|\tilde{V}\tilde{C}) - I(A; B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K})$$

where

$$\mathcal{T} \triangleq H(U|AC) + H(V|BC) - H(UV|ABC)$$

$$\mathcal{T}_1 \triangleq I(AC; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) - I(\tilde{V}\tilde{K}; AC|\tilde{C}\tilde{U})$$

$$\mathcal{T}_2 \triangleq I(BC; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{U}\tilde{K}; BC|\tilde{C}\tilde{V})$$

$$\mathcal{T}_3 \triangleq I(AC; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) - I(\tilde{V}\tilde{K}; AC|\tilde{C}\tilde{U}) + I(B; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}C) - I(\tilde{U}\tilde{K}A; B|C\tilde{C}\tilde{V})$$

$$\mathcal{T}_4 \triangleq I(A; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}C) - I(\tilde{U}\tilde{K}; BC|\tilde{C}\tilde{V}) + I(BC; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{V}\tilde{K}B; A|C\tilde{C}\tilde{U})$$

$$\mathcal{T}_5 \triangleq \frac{1}{2} \left[ I(AC; Y\tilde{Y}\tilde{A}|\tilde{C}\tilde{U}) - I(\tilde{V}\tilde{K}; AC|\tilde{C}\tilde{U}) + I(BC; Z\tilde{Z}\tilde{B}|\tilde{C}\tilde{V}) - I(\tilde{U}\tilde{K}; BC|\tilde{C}\tilde{V}) - I(A; B|C\tilde{C}\tilde{U}\tilde{V}\tilde{K}) \right]$$

*Proof.* This theorem is proved in Section 5.

**Remarks:**

1. The input mapping  $P_{X|ABCUV}$  in the set of distributions  $\mathcal{P}$  can be assumed to be deterministic, i.e.,  $X = f(A, B, C, U, V)$  for some function  $f$ . This is because for a fixed  $P_{ABCUV}$ , optimizing the rate region is equivalent to maximizing a convex functional of  $P_{X|ABCUV}$ . Hence the optimum occurs at one of the corner points, which corresponds to a deterministic  $P_{X|ABCUV}$ .
2. We can recover Marton's achievable rate region for the broadcast channel without feedback by setting  $A = B = \phi$ , and  $C = W$  with  $Q_{C|\tilde{U}\tilde{V}\tilde{K}\tilde{C}} = P_W$ .

### 3 Coding scheme

In this section, we give an informal outline of the proof of Theorem 1. The formal proof is given in Section 5. Let us first consider the case when there is no common message ( $R_0 = 0$ ). Let the message rate pair  $(R_1, R_2)$  lie outside Marton's achievable region [1]. The coding scheme uses a block-Markov superposition strategy, with the communication taking place over  $L$  blocks, each of length  $n$ .

In each block, a fresh pair of messages is encoded using the Marton coding strategy (for the BC without feedback). In block  $l$ , random variables  $U$  and  $V$  carry the fresh information for receivers 1 and 2, respectively. At the end of this block, the receivers are not able to decode the information in  $(U, V)$  completely, so we send

‘resolution’ information in block  $(l + 1)$  using random variables  $(A, B, C)$ . The pair  $(A, C)$  is meant to be decoded by the first receiver, and the pair  $(B, C)$  by the second receiver. Thus in each block, we obtain the channel output by superimposing fresh information on the resolution information for the previous block. At the end of the block, the first receiver decodes  $(A, C)$ , the second receiver decodes  $(B, C)$ , thereby resolving the uncertainty about their messages of the previous block.

*Codebooks:* The  $A$ -,  $B$ -, and  $C$ -codebooks are constructed on the alphabets  $\mathcal{A}$ ,  $\mathcal{B}$ , and  $\mathcal{C}$  respectively. The exact procedure for this construction, and the method for selecting codewords from these codebooks will be described in the sequel. Since  $(A, C)$  is decoded first by receiver 1, conditioned on each codeword pair corresponding to the  $A$ - and  $C$ -codebooks, we construct a  $\mathbf{U}$ -codebook of size  $2^{nR'_1}$  by generating codewords according to  $P_{U|AC}$ . Similarly for each codeword pair in the  $B$ - and  $C$ -codebooks, we construct a  $\mathbf{V}$ -codebook of size  $2^{nR'_2}$  by generating codewords according to  $P_{V|BC}$ . Each  $\mathbf{U}$ -codebook is divided into  $2^{nR_1}$  bins, and each  $\mathbf{V}$ -codebook into  $2^{nR_2}$  bins.

*Encoding:* In each block  $l$ , the encoder chooses a tuple of five codewords  $(\mathbf{A}_l, \mathbf{B}_l, \mathbf{C}_l, \mathbf{U}_l, \mathbf{V}_l)$  as follows. The resolution information for block  $(l - 1)$  is used to select  $(\mathbf{A}_l, \mathbf{B}_l, \mathbf{C}_l)$  from the  $A$ -,  $B$ - and  $C$ -codebooks.  $\mathbf{C}_l$  determines the  $\mathbf{U}$ - and  $\mathbf{V}$ -codebooks to be used to encode the message pair of block  $l$ . Denoting the message pair by  $(m_{1l}, m_{2l})$ , the encoder chooses a  $U$ -codeword from bin  $m_{1l}$  of the  $U$ -codebook and a  $V$ -codeword from bin  $m_{2l}$  of the  $V$ -codebook that are jointly typical according to  $P_{UV|ABC}$ . This pair of jointly typical codewords is set to be  $(\mathbf{U}_l, \mathbf{V}_l)$ .

By standard joint-typicality based covering arguments (see e.g., [16]), this step is successful if the product of the sizes of  $U$ -bin and  $V$ -bin is exponentially larger than  $2^{n(H(U|AC) + H(V|BC) - H(UV|ABC))}$ . Therefore, we have

$$R'_1 + R'_2 - R_1 - R_2 > H(U|AC) + H(V|BC) - H(UV|ABC). \quad (14)$$

These five codewords are combined using the transformation  $P_{X|ABCUV}$  (applied componentwise) to generate the channel input  $\mathbf{X}_l$ .

*Decoding:* After receiving the channel output of block  $l$ , receiver 1 first decodes  $(\mathbf{A}_l, \mathbf{C}_l)$ , and receiver 2 decodes  $(\mathbf{B}_l, \mathbf{C}_l)$ . However, the rates  $R'_1, R'_2$  of the  $U$ - and  $V$ -codebooks are too large for receivers 1 and 2 to uniquely decode  $\mathbf{U}_l$  and  $\mathbf{V}_l$ , respectively. Hence receiver 1 is left with a list of  $U$ -codewords that are jointly typical with its channel output  $\mathbf{Y}_l$  and the just-decoded resolution information  $(\mathbf{A}_l, \mathbf{C}_l)$ ; receiver 2 has a similar list of  $V$ -codewords that are jointly typical with its channel output  $\mathbf{Z}_l$ , and the just-decoded resolution information  $(\mathbf{B}_l, \mathbf{C}_l)$ . The sizes of the lists are nearly equal to  $2^{n(R'_1 - I(U; Y|AC))}$  and  $2^{n(R'_2 - I(V; Z|BC))}$ , respectively. The transmitter receives feedback signal  $\mathbf{S}_l$  in block  $l$ , and resolves these lists in the next block as follows.

In block  $(l + 1)$ , the random variables of block  $l$  are represented using the notation  $\sim$ . Thus we have

$$\tilde{\mathbf{U}}_{l+1} = \mathbf{U}_l, \tilde{\mathbf{V}}_{l+1} = \mathbf{V}_l, \tilde{\mathbf{C}}_{l+1} = \mathbf{C}_l, \tilde{\mathbf{A}}_{l+1} = \mathbf{A}_l, \tilde{\mathbf{B}}_{l+1} = \mathbf{B}_l, \tilde{\mathbf{S}}_{l+1} = \mathbf{S}_l.$$

The random variables  $(U, V, A, B, C, Y, Z, S)$  in block  $l$  are jointly distributed via  $P_{ABC}P_{UV|ABC}P_{YZS|ABCUV}$  chosen from  $\mathcal{P}$  as given in the statement of the theorem.

For block  $l+1$ ,  $(\tilde{\mathbf{U}}_{l+1}, \tilde{\mathbf{V}}_{l+1}) = (\mathbf{U}_l, \mathbf{V}_l)$  can be considered to be a realization of a pair of correlated ‘sources’  $(\tilde{U}$  and  $\tilde{V})$ , jointly distributed according to  $P_{\tilde{U}\tilde{V}|\tilde{S}\tilde{A}\tilde{B}\tilde{C}}$  along with the transmitter side information given by  $(\tilde{\mathbf{A}}_{l+1}, \tilde{\mathbf{B}}_{l+1}, \tilde{\mathbf{S}}_{l+1})$ , and the common side-information  $\tilde{\mathbf{C}}_{l+1}$ . The goal in block  $(l + 1)$  is to transmit this pair of correlated sources over the BC, with



- Receiver 1 needing to decode  $\tilde{\mathbf{U}}_{l+1}$ , treating  $(\tilde{\mathbf{A}}_{l+1}, \tilde{\mathbf{Y}}_{l+1}, \tilde{\mathbf{C}}_{l+1})$  as receiver side-information,
- Receiver 2 needing to decode  $\tilde{\mathbf{V}}_{l+1}$ , treating  $(\tilde{\mathbf{B}}_{l+1}, \tilde{\mathbf{Z}}_{l+1}, \tilde{\mathbf{C}}_{l+1})$  as receiver side-information.

We use the ideas of Han and Costa [8] to transmit this pair of correlated sources over the BC (with appropriate extensions to take into account the different side-information available at the transmitter and the receivers). This is shown in Figure 3. The triplet of correlated random variables  $(A, B, C)$  is used to cover the sources. This triplet carries the resolution information intended to disambiguate the lists of the two receivers. The random variables of block  $(l + 1)$ , given by  $(A, B, C)$  are related to the random variables in block  $l$  via  $Q_{ABC|\tilde{U}\tilde{V}\tilde{C}\tilde{A}\tilde{B}\tilde{S}}$ , chosen from  $\mathcal{Q}$  given in the statement of the theorem. We now describe the construction of the  $A$ -,  $B$ -, and  $C$ - codebooks.

For brevity, we denote the collection of random variables  $(A, B, S)$  as  $K$ , and  $(\mathbf{A}_l, \mathbf{B}_l, \mathbf{S}_l)$  as  $\mathbf{K}_l = \tilde{\mathbf{K}}_{l+1}$ .

*Covering the Sources:* For each  $\tilde{\mathbf{c}} \in \mathcal{C}^n$ , a  $C$ -codebook  $\Psi_C(\tilde{\mathbf{c}})$  of rate  $\rho_0$  is constructed randomly from  $P_{C|\tilde{C}}$ . For every realization of  $\tilde{\mathbf{u}} \in \mathcal{U}^n$ ,  $\tilde{\mathbf{c}} \in \mathcal{C}^n$ , and  $\mathbf{c} \in \mathcal{C}^n$ , an  $A$ -codebook  $\Psi_A(\tilde{\mathbf{u}}, \tilde{\mathbf{c}}, \mathbf{c})$  of rate  $\rho_1$  is constructed with codewords picked randomly according to  $P_{A|\tilde{U}\tilde{C}C}$ . (In Figure 3, we see that in addition to  $\tilde{C}$ , receiver 1 also has  $(\tilde{A}, \tilde{Y})$  as side-information. However, we do not pick the  $A$ -codebook conditioned on these random variables since they are not available at all three terminals.) For every realization of  $\tilde{\mathbf{v}} \in \mathcal{V}^n$ ,  $\tilde{\mathbf{c}} \in \mathcal{C}^n$ , and  $\mathbf{c} \in \mathcal{C}^n$ , a  $B$ -codebook  $\Psi_B(\tilde{\mathbf{v}}, \tilde{\mathbf{c}}, \mathbf{c})$  of rate  $\rho_2$  is constructed with codewords picked randomly according to  $P_{B|\tilde{V}\tilde{C}C}$ .

At the beginning of block  $(l + 1)$ , for a given realization  $(\tilde{\mathbf{U}}_{l+1}, \tilde{\mathbf{V}}_{l+1}, \tilde{\mathbf{K}}_{l+1}, \tilde{\mathbf{C}}_{l+1})$ , of correlated ‘sources’, and side information, the encoder chooses a triplet of codewords  $(\mathbf{A}_{l+1}, \mathbf{B}_{l+1}, \mathbf{C}_{l+1})$  from the appropriate  $A$ -,  $B$ - and  $C$ -codebooks such that the two tuples are jointly typical according to  $P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}}Q_{ABC|\tilde{U}\tilde{K}\tilde{V}\tilde{C}}$ . The channel input  $\mathbf{X}_{l+1}$  is generated by fusing this  $(\mathbf{A}_{l+1}, \mathbf{B}_{l+1}, \mathbf{C}_{l+1})$  with the pair of codewords  $(\mathbf{U}_{l+1}, \mathbf{V}_{l+1})$ , which carry fresh information in block  $(l + 1)$ .

Now consider the general case when  $R_0 > 0$ . We can use the random variable  $C$  to encode common information to be decoded by both receivers. Hence  $C$  serves two purposes: it is used to (a) cover the correlated sources and transmitter side-information and is thus part of the resolution information, and (b) to carry fresh information that is decoded by both receivers. We note that in every block, two communication tasks are being accomplished simultaneously. The first is joint source-channel coding of correlated sources over the BC, accomplished via  $(A, B, C)$ ; the second is Marton coding of the fresh information, accomplished via  $(U, V, C)$ <sup>1</sup>.  $C$  can be made to assume the dual role of the common random variable associated with both these tasks.

*Analysis:* For this encoding to be successful, we need the following covering conditions. These are the same conditions that appear in the Han-Costa scheme (see [17, Lemma 14.1]), with  $(\tilde{U}, \tilde{K})$  and  $(\tilde{V}, \tilde{K})$  assuming

<sup>1</sup>Recall that in Marton’s achievable region for the BC without feedback, there is a random variable  $W$  meant to be decoded by both receivers.

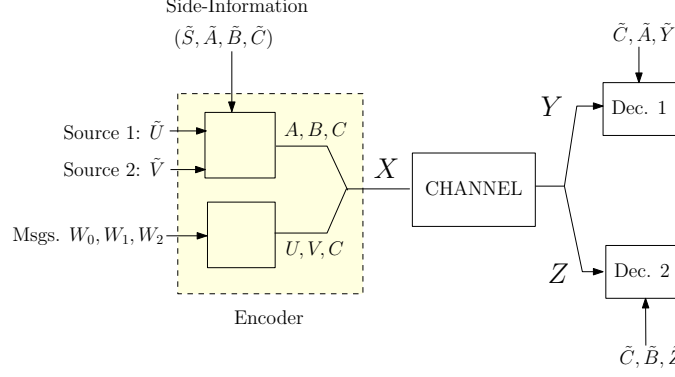


Figure 3: Transmitting correlated sources with side-information at the receivers through  $(A, B, C)$ , and fresh information through  $U, V, C$ .  $C$  plays the dual role - it is used to cover the correlated sources as well as carry fresh information.

the roles of the two sources being covered.<sup>2</sup>

$$\rho_0 > I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (15)$$

$$\rho_0 + \rho_1 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (16)$$

$$\rho_0 + \rho_2 > I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (17)$$

$$\rho_0 + \rho_1 + \rho_2 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(A; B|\tilde{U}\tilde{K}\tilde{V}C\tilde{C}) + I(\tilde{U}\tilde{K}\tilde{V}; C|\tilde{C}) + R_0 \quad (18)$$

At the end of block  $(l+1)$ , receiver 1 determines  $\mathbf{U}_l = \tilde{\mathbf{U}}_{l+1}$  by finding the pair  $(\tilde{\mathbf{U}}_{l+1}, \mathbf{A}_{l+1}, \mathbf{C}_{l+1})$  using joint typical decoding in the composite  $U$ -,  $A$ -, and  $C$ -codebooks. A similar procedure is followed at the second receiver. For the decoding to be successful, we need the following packing conditions.

$$R'_1 + \rho_0 + \rho_1 < I(\tilde{U}; Y\tilde{Y}\tilde{A}|\tilde{C}) + I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C}) \quad (19)$$

$$R'_1 + \rho_1 < I(\tilde{U}; Y\tilde{A}\tilde{Y}C|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C}) \quad (20)$$

$$R'_2 + \rho_0 + \rho_2 < I(\tilde{V}; Z\tilde{Z}\tilde{B}|\tilde{C}) + I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}C\tilde{C}) \quad (21)$$

$$R'_2 + \rho_2 < I(\tilde{V}; Z\tilde{B}\tilde{Z}C|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}C\tilde{C}) \quad (22)$$

$$\rho_0 + \rho_1 < I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C}) \quad (23)$$

$$\rho_0 + \rho_2 < I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}C\tilde{C}) \quad (24)$$

$$\rho_1 < I(A; Y\tilde{A}\tilde{Y}|\tilde{U}C\tilde{C}) \quad (25)$$

$$\rho_2 < I(B; Z\tilde{B}\tilde{Z}|\tilde{V}C\tilde{C}) \quad (26)$$

Performing Fourier-Motzkin elimination on equations (14), (15-18) and (19-26), we obtain the statement of the theorem.

To get a single-letter characterization of achievable rates, we need to ensure that the random variables in each block follow a stationary joint distribution. We now describe how we ensure that the sequences in each block are jointly distributed according to

$$P_{ABC} \cdot P_{UV|ABC} \cdot P_{X|ABCUV} \cdot P_{YZS|X} \quad (27)$$

<sup>2</sup>Though  $\tilde{K} = (\tilde{A}, \tilde{B}, \tilde{S})$  is included in the covering, it is not required to be explicitly decoded at either receiver.

for some chosen  $P_{ABC}$ ,  $P_{UV|ABC}$ , and  $P_{X|ABCUV}$ .

Suppose that the sequences in a given block are jointly distributed according to (27). In the next block, these sequences become the source pair  $(\tilde{U}, \tilde{V})$ , transmitter side-information  $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{S})$  and the side information at the two receivers –  $(\tilde{C}, \tilde{A}, \tilde{Y})$  and  $(\tilde{C}, \tilde{B}, \tilde{Z})$ , respectively. To cover the source pair with  $(A, B, C)$ , we pick a conditional distribution  $Q_{ABC|\tilde{A}\tilde{B}\tilde{C}\tilde{U}\tilde{V}\tilde{S}}$  such that the covering sequences are distributed according to  $P_{ABC}$ . This holds when the consistency condition given by (6) is satisfied. We thereby ensure that the sequences in each block are jointly distributed according to (27). Our technique of exploiting the correlation induced by feedback is similar in spirit to the coding scheme of Han for two-way channels [18].

We note that the transmitter side information  $\tilde{K} = (\tilde{A}\tilde{B}\tilde{S})$  is exploited at the encoder in the covering operation implicitly, without using codebooks conditioned on  $\tilde{K}$ . This is because this side information is only partially available at the receivers, with receiver 1 having only  $(\tilde{A}, \tilde{Y})$ , and receiver 2 having only  $(\tilde{B}, \tilde{Z})$ . Hence the coding approach does not depend on any assumptions on the nature of the generalized feedback signal  $S$ . This is in contrast to communication over a multiple-access channel with feedback, where there is a significant difference between noiseless feedback and noisy feedback [19].

## 4 Special Cases and Examples

In this section, we obtain a simpler version of the rate region of Theorem 1 and use it to compute achievable rates for a few examples.

### 4.1 A Simpler Rate Region

**Corollary 4.1.** *Given a broadcast channel with generalized feedback  $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}, P_{YZS|X})$ , define any joint distribution  $P$  of the form*

$$P_{C_0} P_{WUV} P_{X|WUVC_0} P_{YZS|X}. \quad (28)$$

*for some discrete random variables  $W, U, V, C_0$ . Let  $(C_0, W, U, V, X, Y, Z, S)$  and  $(\tilde{C}_0, \tilde{W}, \tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{S})$  be two sets of variables each distributed according to  $P$  and jointly distributed as*

$$P_{\tilde{C}_0 \tilde{W} \tilde{U} \tilde{V} \tilde{X} \tilde{Y} \tilde{Z} \tilde{S}} Q_{C_0|\tilde{C}_0 \tilde{W} \tilde{U} \tilde{V} \tilde{S}} P_{WUVXYZS|C_0}. \quad (29)$$

*where  $Q_{C_0|\tilde{C}_0 \tilde{W} \tilde{U} \tilde{V} \tilde{S}}$  is a distribution such that*

$$P_{C_0}(c_0) = \sum_{\tilde{c}_0, \tilde{w}, \tilde{u}, \tilde{v}, \tilde{s}} Q_{C_0|\tilde{C}_0 \tilde{W} \tilde{U} \tilde{V} \tilde{S}}(c_0|\tilde{c}_0, \tilde{w}, \tilde{u}, \tilde{v}, \tilde{s}) P(\tilde{c}_0, \tilde{w}, \tilde{u}, \tilde{v}, \tilde{s}), \quad \forall c_0 \in \mathcal{C}_0. \quad (30)$$

Then the following region is achievable.

$$R_0 < \min\{\mathcal{T}_1, \mathcal{T}_2\} \quad (31)$$

$$R_0 + R_1 < I(UW; Y|C_0) + I(C_0; Y|\tilde{Y}\tilde{C}_0\tilde{W}) + I(C_0; \tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{U}) \quad (32)$$

$$R_0 + R_2 < I(VW; Z|C_0) + I(C_0; Z|\tilde{Z}\tilde{C}_0\tilde{W}) + I(C_0; \tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{V}) \quad (33)$$

$$R_0 + R_1 + R_2 < I(UW; Y|C_0) + I(C_0; Y|\tilde{Y}\tilde{C}_0\tilde{W}) + I(C_0; \tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{U}) \\ + I(C_0\tilde{Z}; \tilde{V}|\tilde{C}_0\tilde{W}) - I(U; V|W) \quad (34)$$

$$R_0 + R_1 + R_2 < I(VW; Z|C_0) + I(C_0; Z|\tilde{Z}\tilde{C}_0\tilde{W}) + I(C_0; \tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{V}) \\ + I(C_0\tilde{Y}; \tilde{U}|\tilde{C}_0\tilde{W}) - I(U; V|W) \quad (35)$$

$$2R_0 + R_1 + R_2 < I(UW; Y|C_0) + I(C_0; Y|\tilde{Y}\tilde{C}_0\tilde{W}) + I(C_0; \tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{U}) \\ + I(VW; Z|C_0) + I(C_0; Z|\tilde{Z}\tilde{C}_0\tilde{W}) + I(C_0; \tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{V}) - I(U; V|W) \quad (36)$$

where

$$\mathcal{T}_1 \triangleq I(C_0; \tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) + I(C_0W; Y|\tilde{Y}\tilde{C}_0\tilde{W}\tilde{U}) - I(\tilde{V}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{U})$$

$$\mathcal{T}_2 \triangleq I(C_0; \tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) + I(C_0W; Z|\tilde{Z}\tilde{C}_0\tilde{W}\tilde{V}) - I(\tilde{U}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{V})$$

*Proof.* In Theorem 1, set  $A = B = \phi$ , and  $C = (C_0, W)$ , with

$$Q_{C|\tilde{C}\tilde{U}\tilde{V}\tilde{S}} = Q_{C_0W|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{S}} = P_W Q_{C_0|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{S}}.$$

For this choice, we have  $Q_{C|\tilde{C}\tilde{U}\tilde{V}\tilde{S}} \in \mathcal{Q}(P)$  if (30) is satisfied.  $\square$

## 4.2 The AWGN Broadcast Channel with Noisy Feedback

We now use Corollary 4.1 to compute achievable rates for the scalar AWGN broadcast channel with noisy feedback from one receiver. We compare the obtained sum rate with: a) the maximum sum rate in the absence of feedback, b) the achievable region of Bhaskaran [6] for the case with noiseless feedback from one receiver, and c) the achievable region of Ozarow and Leung [5] for noiseless feedback from both receivers. We note that the coding schemes in both [6] and [5] are linear schemes based on Schalkwijk-Kailath coding for the AWGN channel [20], and cannot be used when there is noise in the feedback link [21]. Our rate region also includes the possibility of a common message to both receivers. The coding schemes of [5] and [6] are constructed only for private messages.

The channel, with  $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$ , is described by

$$Y = X + N_1, \quad Z = X + N_2, \quad (37)$$

where  $N_1, N_2$  are Gaussian noise variables with zero mean and covariance matrix

$$K_{N_1, N_2} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

where  $\rho \in [-1, 1]$ .  $N_1$  and  $N_2$  are independent of the channel input  $X$ . The input sequence  $\mathbf{x}$  for each block

satisfies an average power constraint  $\sum_{i=1}^n x_i^2 \leq nP$ .

In the absence of feedback, the capacity region of the AWGN broadcast channel is known [22, 23] and can be obtained from Marton's inner bound using the following choice of random variables.

$$V = \sqrt{\alpha P} Q_2, \quad U = \sqrt{\alpha P} Q_1 + \frac{\alpha P}{\alpha P + \sigma^2} V$$

where  $\alpha \in (0, 1)$ , and  $Q_1, Q_2$  are independent  $\mathcal{N}(0, 1)$  random variables. The Marton sum rate is then given by

$$R_{\text{no-FB}} = I(V; Z) + I(U; Y) - I(U; V) = \frac{1}{2} \log_2 \left( 1 + \frac{P}{\sigma^2} \right). \quad (38)$$

This is essentially the 'writing on dirty paper' coding strategy [24, 25]: for the channel from  $U$  to  $Y$ ,  $V$  can be considered as channel state information known at the encoder. We note that an alternate way of achieving the no-feedback capacity region of the AWGN broadcast channel is through superposition coding [2].<sup>3</sup>

Using Corollary 4.1, we now compute an achievable region for the channel (37) with noisy feedback from transmitter 1 alone. The feedback signal is given by

$$S = Y + N_f \quad (39)$$

where  $N_f$  is additive white Gaussian noise on the feedback link distributed as  $\mathcal{N}(0, \sigma_f^2)$ .  $N_f$  is independent of  $X, Y, Z, N_1$  and  $N_2$ .

To motivate the choice of joint distribution, let us first consider the case of noiseless feedback, i.e.,  $N_f = 0$ .

*Noiseless Feedback:* The joint distribution  $P_{C_0} P_{UV} P_{X|C_0 UV}$  is chosen as

$$V = \sqrt{\alpha P_1} Q_2, \quad U = \sqrt{\alpha P_1} Q_1 + \beta V \quad (40)$$

$$X = \sqrt{P - P_1} C_0 + \sqrt{\alpha P_1} Q_2 + \sqrt{\alpha P_1} Q_1 \quad (41)$$

where  $Q_1, Q_2, C_0$  are independent Gaussians with zero mean and unit variance and  $\alpha, \beta \in (0, 1)$ ,  $P_1 \in (0, P)$  are parameters to be optimized later.

Next we define a conditional distribution  $Q_{C_0|\tilde{C}_0 \tilde{U} \tilde{V} \tilde{Y} \tilde{Z}}$  that satisfies (30). Let

$$\tilde{T}_1 = \frac{\tilde{U} - E[\tilde{U}|\tilde{Y} \tilde{C}_0]}{\sqrt{E[(\tilde{U} - E[\tilde{U}|\tilde{Y} \tilde{C}_0])^2]}}. \quad (42)$$

Then define  $Q_{C_0|\tilde{C}_0 \tilde{U} \tilde{V} \tilde{Y} \tilde{Z}}$  by the relation

$$C_0 = \sqrt{1 - D} \tilde{T}_1 + \zeta \quad (43)$$

where  $\zeta$  is a  $\mathcal{N}(0, D)$  random variable independent of  $(\tilde{C}_0, \tilde{U}, \tilde{V}, \tilde{Y}, \tilde{Z})$ .

In words,  $\tilde{T}_1$  is the normalized error in the estimate of  $\tilde{U}$  at receiver 1. This estimation error is quantized at distortion level  $D$  and suitably scaled to obtain  $C_0$ . Thus, in each block,  $C_0$  represents a quantized version of the estimation error at receiver 1 in the previous block. If we similarly denote by  $\tilde{T}_2$  the error in the estimate of  $\tilde{V}$  at receiver 2, then  $\tilde{T}_2$  is correlated with  $\tilde{T}_1$ . Therefore,  $C_0$  simultaneously plays the role of conveying

---

<sup>3</sup>Theorem 1 was established for a discrete memoryless broadcast channel with feedback. These theorems can be extended to the AWGN broadcast channel using a similar proof, recognizing that in the Gaussian case superposition is equivalent to addition.

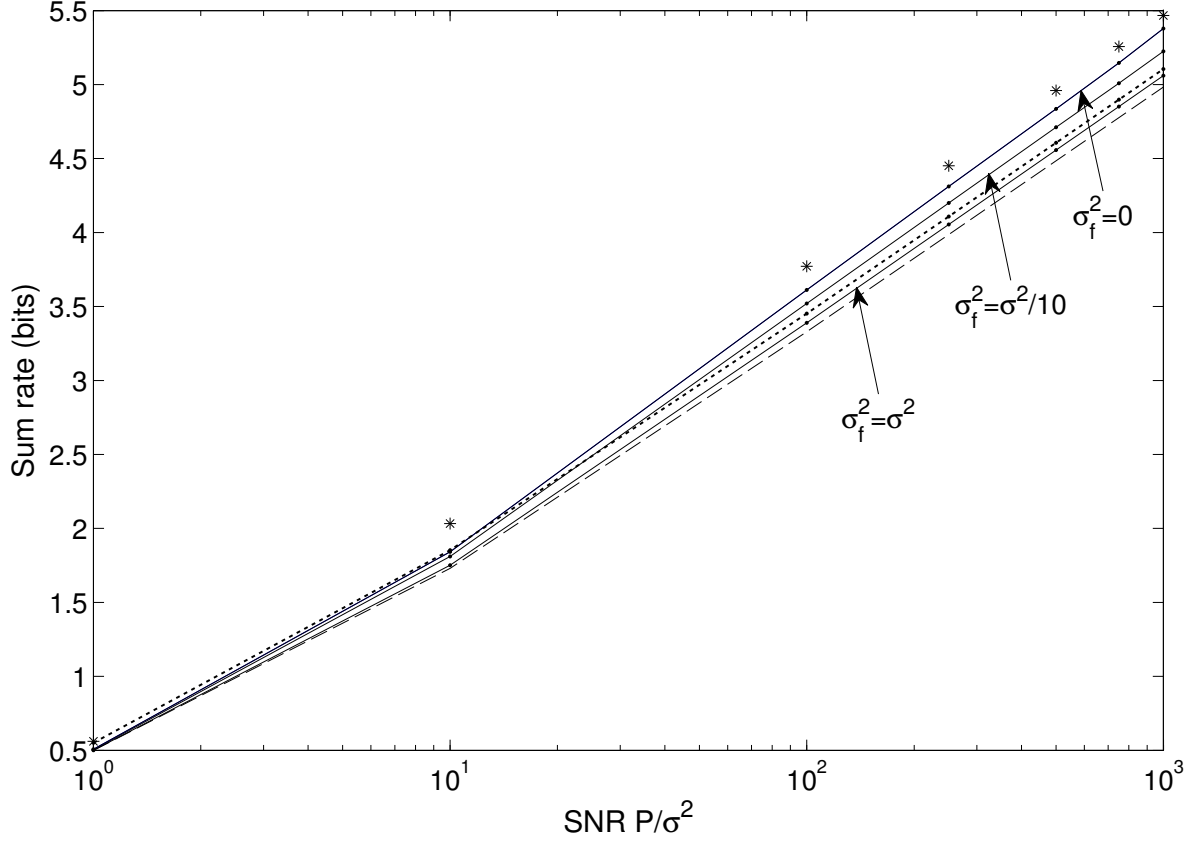


Figure 4: Achievable sum rates for the AWGN broadcast channel with noisy feedback from receiver 1. Noise correlation  $\rho = 0$ . The three solid lines show the sum-rates computed using Corollary 4.1 for feedback noise variance  $\sigma_f^2 = 0, \sigma^2/10$ , and  $\sigma^2$ . The dashed line at the bottom is the no-feedback sum rate, the dotted line in the middle is the sum-rate of the Bhaskaran scheme, and the \* symbols at the top are the sum rate of the Ozarow-Leung scheme.

information about  $\tilde{T}_1$  to receiver 1, and about  $\tilde{T}_2$  to receiver 2. With this choice of joint distribution, the information quantities in Corollary 4.1 can be computed.

*Noisy Feedback:* When the feedback is noisy, the transmitter does not know  $\tilde{Y}$ , and so cannot compute  $\tilde{U} - E[\tilde{U}|\tilde{Y}\tilde{C}_0]$  in (42) which was used to generate  $C_0$ . Instead, the transmitter can compute an *estimate* of the error at receiver 1. We now define  $\tilde{T}_1$  as

$$\tilde{T}_1 = \frac{\Delta}{\sqrt{E[\Delta^2]}} \quad (44)$$

where

$$\begin{aligned} \Delta &= E \left[ (\tilde{U} - E[\tilde{U}|\tilde{Y}\tilde{C}_0]) \mid \tilde{U}\tilde{V}\tilde{C}_0\tilde{S} \right] \\ &= \sqrt{\alpha P_1} \frac{\sigma^2 + \bar{\alpha}\bar{\beta}P_1}{P_1 + \sigma^2} \tilde{Q}_1 + \sqrt{\bar{\alpha}P_1} \frac{\beta\sigma^2 - \alpha\bar{\beta}P_1}{P_1 + \sigma^2} \tilde{Q}_2 - \frac{P_1(\alpha + \beta\bar{\alpha})}{P_1 + \sigma^2} \frac{\sigma^2}{\sigma^2 + \sigma_f^2} (\tilde{S} - \tilde{X}). \end{aligned} \quad (45)$$

As before,  $C_0$  is defined by (43) with  $\tilde{T}_1$  given by (44), and the input  $X$  is defined by (41). With this choice

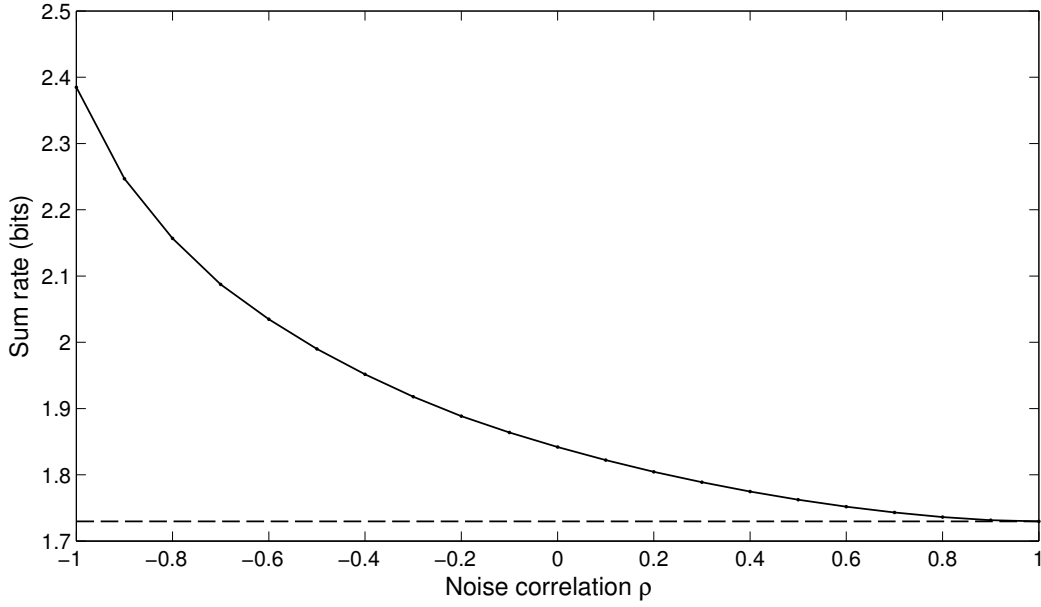


Figure 5: Variation of the sum-rate vs correlation coefficient of  $(N_1, N_2)$ .  $P/\sigma^2 = 10$  and there is noiseless feedback from receiver 1. The dashed line shows the no-feedback sum-rate.

of joint distribution, the information quantities required to evaluate Corollary 4.1 are computed and listed in Appendix A.

For different values of the signal-to-noise ratio  $P/\sigma^2$ , feedback noise variance  $\sigma_f^2$  and correlation coefficient  $\rho$ , we can compute the maximum sum rate by numerically optimizing over the parameters  $(\alpha, \beta, D, P_1)$ . For the case where the noises at the two receivers are independent ( $\rho = 0$ ), the maximum sum rate is plotted in Figure 4 for  $\sigma_f^2 = \sigma^2$ ,  $\frac{\sigma^2}{10}$  and 0 ( $\sigma_f^2 = 0$  is noiseless feedback). The figure also shows the sum rate in the absence of feedback, the sum rate of the Bhaskaran scheme [6] for noiseless feedback from one receiver, and the maximum sum rate of the Ozarow-Leung scheme with noiseless feedback from both receivers.

We see that the obtained sum rate is higher than the no-feedback sum rate even with feedback noise variance  $\sigma_f^2 = \sigma^2$ , and increases as  $\sigma_f^2$  decreases. We also observe that for  $\sigma_f^2 = 0$  (noiseless feedback), the sum rate of the proposed rate-region is higher than the Bhaskaran sum rate for high SNR. Concretely, for  $P/\sigma^2 = 10, 100$  and 1000, our region yields sum rates of 1.842, 3.612 and 5.378, respectively; the Bhaskaran sum rates for these SNR values are 1.852, 3.452 and 5.105. The Ozarow-Leung scheme yields higher sum rates than the proposed region, but we emphasize that it uses noiseless feedback from both receivers. Another difference is that both the Ozarow-Leung and Bhaskaran schemes are specific to the AWGN broadcast channel and do not extend to other discrete memoryless broadcast channels, unlike the scheme in this paper.

Figure 5 shows the effect of  $\rho$  (the correlation coefficient of  $N_1, N_2$ ) on the noiseless feedback sum-rate with  $P/\sigma^2$  held fixed. Note that the sum-rate without feedback does not change with  $\rho$  as long as the individual noise variances remain unchanged [2]. We observe that the sum-rate decreases monotonically with the noise correlation and is equal to the no-feedback rate at  $\rho = 1$ . This is consistent with the fact that feedback does not increase the capacity of the AWGN broadcast channel with  $\rho = 1$  since it is physically degraded (in fact, we effectively have a point-to-point channel when  $\rho = 1$ ).

### 4.3 Comparison with the Shayevitz-Wigger (S-W) Rate Region

An achievable rate region for the broadcast channel with feedback was independently proposed by Shayevitz and Wigger [15]. Their coding scheme can be summarized as follows. In the first block, the encoder transmits at rates outside than the Marton region. The receivers cannot decode, and as discussed earlier, the information needed to resolve the ambiguity at the two receivers is correlated. This resolution information is transmitted in the next block through separate source and channel coding. The correlated resolution information is first quantized into three parts: a common part, and a private part for each receiver. This quantization is performed using a generalization of Gray-Wyner coding [26]. The quantization indices representing the correlated information are then transmitted together with fresh information for the second block using Marton coding.

While the S-W scheme is also a block-Markov superposition scheme with the Marton coding as the starting point, the S-W scheme differs from the one proposed in this paper in two aspects:

1. Separate source and channel coding
2. Backward decoding

While separate source and channel coding can be considered a special case of joint source-channel coding, the backward decoding technique in [15] uses the resolution information in a different way than our scheme. In particular, the covering random variables in each block are decoded first and serve as extra ‘outputs’ at the receivers that augment the channel outputs. This difference in the decoding strategy makes a general comparison of the two rate regions difficult.

In Appendix B, we show that the class of valid joint distributions for the S-W region can be obtained using our coding scheme via a specific choice of the covering variables  $(A, B, C)$ . The rate region of Theorem 1 evaluated with this class of distributions is given in (81). We observe that the bounds on  $R_0 + R_1$ ,  $R_0 + R_2$ ,  $R_0 + R_1 + R_2$  and  $2R_0 + R_1 + R_2$  are larger than the corresponding bounds in the S-W region. However, our region has an additional  $R_0$  constraint which is not subsumed by the other constraints. Therefore a general statement about the inclusion of one region in the other does not seem possible. In the following, we focus on the two examples discussed in [15] and show that the feedback rates of the S-W region can also be obtained using Corollary 4.1.

#### The Generalized Dueck Broadcast Channel

This is a generalization of the Dueck example discussed in Section 1. The input  $X$  is a binary triple  $(X_0, X_1, X_2)$ . The output of the two receivers 1 are  $Y = (X_0 + N_0, X_1 + N_1)$  and  $Z = (X_0 + N_0, X_1 + N_2)$  where  $(N_0, N_1, N_2)$  are binary random variables with distribution  $P_{N_0, N_1, N_2}$  such that

$$H(N_0, N_1) \leq 1, \quad H(N_0, N_2) \leq 1.$$

We evaluate the rate region of Corollary 4.1 for noiseless feedback from receiver 1 with the following joint distribution.

$$\begin{aligned} (W, U, V) &\sim P_W P_U P_V \text{ with } P_W, P_U, P_V \sim \text{Bernoulli}\left(\frac{1}{2}\right), \\ Q_{C_0|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{Y}} : C_0 &= \tilde{Y} \oplus \tilde{U} = \tilde{N}_1, \\ X : (X_0, X_1, X_2) &= (W, U, V) \end{aligned} \tag{46}$$



With this choice of  $Q$ ,  $C_0$  is a Bernoulli random variable with the same distribution as  $N_1$ . With the joint distribution above, the mutual information quantities in Corollary 4.1 can be computed to be

$$\begin{aligned} I(UW; Y|C_0) &= 2 - H(N_0, N_1), & I(VW; Z|C_0) &= 2 - H(N_0, N_2), & I(C_0; Y|\tilde{Y}\tilde{C}_0\tilde{W}) &= I(C_0; Z|\tilde{Z}\tilde{C}_0\tilde{W}) = 0, \\ I(C_0; \tilde{Y}|\tilde{C}_0\tilde{W}\tilde{U}) &= H(N_1), & I(C_0; \tilde{Z}|\tilde{C}_0\tilde{W}\tilde{V}) &= H(N_1) - H(N_1|N_0, N_2), & I(C_0\tilde{Y}; \tilde{U}|\tilde{C}_0\tilde{W}) &= 1, \\ I(C_0\tilde{Z}; \tilde{V}|\tilde{C}_0\tilde{W}) &= 1 - H(N_2|N_0N_1), & I(\tilde{V}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{U}) &= I(\tilde{U}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{V}) = H(N_1), \\ I(C_0W; Y|\tilde{Y}\tilde{C}_0\tilde{W}\tilde{U}) &= I(C_0W; Z|\tilde{Z}\tilde{C}_0\tilde{W}\tilde{V}) = 1 - H(N_0). \end{aligned}$$

The rate region is given by

$$\begin{aligned} R_0 &\leq 1 - H(N_0) - H(N_1|N_0, N_2) \\ R_0 + R_1 &\leq 2 - H(N_0, N_1) \\ R_0 + R_2 &\leq 2 - H(N_0, N_1, N_2) \\ R_0 + R_1 + R_2 &\leq 3 - H(N_0, N_1, N_2) \end{aligned} \tag{47}$$

The roles of  $R_1, R_2$  in (47) can be exchanged by choosing  $C_0 = \tilde{Z} \oplus \tilde{V} = \tilde{N}_2$ . Thus the following feedback capacity region obtained in [15] is achievable.

$$R_1 \leq 2 - H(N_0, N_1), \quad R_2 \leq 2 - H(N_0, N_2), \quad R_1 + R_2 \leq 3 - H(N_0, N_1, N_2).$$

### The Noisy Blackwell Broadcast Channel

This generalization of the Blackwell channel has ternary input alphabet  $\mathcal{X} = \{0, 1, 2\}$ , binary output alphabets  $\mathcal{Y} = \mathcal{Z} = \{0, 1\}$  and channel law given by

$$Y = \begin{cases} N & X = 0 \\ 1 - N & X = 1, 2 \end{cases} \quad Z = \begin{cases} N & X = 0, 1 \\ 1 - N & X = 2 \end{cases}$$

where  $N \sim \text{Bernoulli}(p)$  is a noise variable independent of  $X$ . With noiseless feedback from both receivers, the rate region obtained in [15] can also be obtained using Corollary 4.1 with the following joint distribution.

$$\begin{aligned} P_W(0) &= P_W(1) = \frac{1}{2}, \\ P_{UV|W}(0, 0|W = 0) &= \alpha, \quad P_{UV|W}(1, 1|W = 0) = \beta, \quad P_{UV|W}(1, 0|W = 0) = 1 - \alpha - \beta, \\ P_{UV|W}(0, 0|W = 1) &= \beta, \quad P_{UV|W}(1, 1|W = 1) = \alpha, \quad P_{UV|W}(1, 0|W = 1) = 1 - \alpha - \beta, \\ X &= U + V, \\ Q_{C_0|\tilde{C}_0\tilde{W}\tilde{U}\tilde{V}\tilde{Y}} : C_0 &= \tilde{Y} \oplus \tilde{U} = \tilde{Z} \oplus \tilde{V} = \tilde{N}. \end{aligned} \tag{48}$$

With  $h(\cdot)$  denoting the binary entropy function and  $x \star y = x(1 - y) + y(1 - x)$ , the mutual information

quantities in Corollary 4.1 are

$$\begin{aligned}
I(UW; Y|C_0) &= I(VW; Z|C_0) = h\left(p \star \frac{\alpha + \beta}{2}\right) - h(p) \\
I(C_0; Y|\tilde{Y}\tilde{C}_0\tilde{W}) &= I(C_0; Z|\tilde{Z}\tilde{C}_0\tilde{W}) = 0 \\
I(\tilde{V}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{U}) &= I(\tilde{U}\tilde{S}; C_0|\tilde{C}_0\tilde{W}\tilde{V}) = h(p) \\
I(C_0\tilde{Z}; \tilde{V}|\tilde{C}_0\tilde{W}) - I(U; V|W) &= H(V|UW) = \frac{1}{2} \left( \bar{\beta}h\left(\frac{\alpha}{\bar{\beta}}\right) + \bar{\alpha}h\left(\frac{\beta}{\bar{\alpha}}\right) \right) \\
I(C_0\tilde{Y}; \tilde{U}|\tilde{C}_0\tilde{W}) - I(U; V|W) &= H(U|VW) = \frac{1}{2} \left( \bar{\beta}h\left(\frac{\alpha}{\bar{\beta}}\right) + \bar{\alpha}h\left(\frac{\beta}{\bar{\alpha}}\right) \right) \\
I(W; Y|\tilde{Y}\tilde{C}_0\tilde{W}\tilde{U}) &= I(W; Z|\tilde{Z}\tilde{C}_0\tilde{W}\tilde{U}) = h\left(p \star \frac{\alpha + \beta}{2}\right) - \frac{1}{2}h\left(\frac{\alpha p + \bar{\alpha}\bar{p}}{2}\right) - \frac{1}{2}h\left(\frac{\beta p + \bar{\beta}\bar{p}}{2}\right)
\end{aligned}$$

The rate region is then given by

$$\begin{aligned}
R_0 &\leq h\left(p \star \frac{\alpha + \beta}{2}\right) - \frac{1}{2}h\left(\frac{\alpha p + \bar{\alpha}\bar{p}}{2}\right) - \frac{1}{2}h\left(\frac{\beta p + \bar{\beta}\bar{p}}{2}\right) \\
R_0 + R_1 &\leq h\left(p \star \frac{\alpha + \beta}{2}\right) - h(p) \\
R_0 + R_2 &\leq h\left(p \star \frac{\alpha + \beta}{2}\right) - h(p) \\
R_0 + R_1 + R_2 &\leq h\left(p \star \frac{\alpha + \beta}{2}\right) + \frac{1}{2} \left( \bar{\beta}h\left(\frac{\alpha}{\bar{\beta}}\right) + \bar{\alpha}h\left(\frac{\beta}{\bar{\alpha}}\right) \right) - h(p)
\end{aligned} \tag{49}$$

For  $R_0 = 0$ , this matches the rate-region obtained in [15] for this channel.

## 5 Proof of Theorem 1

### 5.1 Preliminaries

We shall use the notion of typicality as defined in [17, 27]. Consider finite sets  $\mathcal{Z}_1, \mathcal{Z}_2$  and any distribution  $P_{\mathcal{Z}_1\mathcal{Z}_2}$  on them.

**Definition 5.1.** For any  $\epsilon > 0$ , the set of jointly  $\epsilon$ -typical sequences with respect to  $P_{\mathcal{Z}_1\mathcal{Z}_2}$  is defined as

$$\mathcal{A}_\epsilon^{(n)}(P_{\mathcal{Z}_1\mathcal{Z}_2}) = \left\{ (\mathbf{z}_1, \mathbf{z}_2) : \left| \frac{1}{n}N(a, b | \mathbf{z}_1, \mathbf{z}_2) - P_{\mathcal{Z}_1\mathcal{Z}_2}(a, b) \right| \leq \epsilon P_{\mathcal{Z}_1\mathcal{Z}_2}(a, b), \text{ for all } (a, b) \in \mathcal{Z}_1 \times \mathcal{Z}_2 \right\}$$

where  $N(a, b | \mathbf{z}_1, \mathbf{z}_2)$  is the number of occurrences of the symbol pair  $(a, b)$  in the sequence pair  $(\mathbf{z}_1, \mathbf{z}_2)$ . For any  $\mathbf{z}_1 \in \mathcal{Z}_1^n$ , define the set of conditionally  $\epsilon$ -typical sequences as

$$\mathcal{A}_\epsilon^n(\mathcal{Z}_2|\mathbf{z}_1) = \{\mathbf{z}_2 : (\mathbf{z}_1, \mathbf{z}_2) \in \mathcal{A}_\epsilon^{(n)}(P_{\mathcal{Z}_1\mathcal{Z}_2})\}.$$

The following are some basic properties of typical sequences that will be used in the proof.  $\delta(\epsilon)$  will be used to denote a generic positive function of  $\epsilon$  that tends to zero as  $\epsilon \rightarrow 0$ .

**Property 0:** For all  $\epsilon > 0$ , and for all sufficiently large  $n$ , we have  $P_{\mathcal{Z}_1, \mathcal{Z}_2}^n[\mathcal{A}_\epsilon^{(n)}(P_{\mathcal{Z}_1, \mathcal{Z}_2})] > 1 - \epsilon$ .

**Property 1:** Let  $\mathbf{z}_1 \in \mathcal{A}_\epsilon^{(n)}(P_{Z_1})$  for some  $\epsilon > 0$ . If  $\mathbf{Z}_2$  is generated according to the product distribution  $\prod_{i=1}^n P_{Z_2|Z_1}(\cdot|z_{1i})$ , then for all  $\epsilon' > \epsilon$

$$\lim_{n \rightarrow \infty} \Pr[(\mathbf{z}_1, \mathbf{Z}_2) \in \mathcal{A}_{\epsilon'}^{(n)}(P_{Z_1 Z_2})] = 1.$$

**Property 2:** For every  $\mathbf{z}_1 \in \mathcal{Z}_1^n$ , the size of the conditionally  $\epsilon$ -typical set is upper bounded as

$$|\mathcal{A}_\epsilon^n(Z_2|\mathbf{z}_1)| \leq 2^{n(H(Z_2|Z_1) + \delta(\epsilon))}.$$

If  $\mathbf{z}_1 \in \mathcal{A}_\epsilon^n(P_{Z_1})$ , then for any  $\epsilon' > \epsilon$  and  $n$  sufficiently large

$$|\mathcal{A}_\epsilon^n(Z_2|\mathbf{z}_1)| \geq 2^{n(H(Z_2|Z_1) - \delta(\epsilon'))}.$$

**Property 3:** If  $(\mathbf{z}_1, \mathbf{z}_2) \in \mathcal{A}_\epsilon^n(P_{Z_1, Z_2})$ , then

$$2^{-n(H(Z_2|Z_1) + \delta(\epsilon))} \leq P_{Z_2|Z_1}(\mathbf{z}_2|\mathbf{z}_1) \leq 2^{-n(H(Z_2|Z_1) - \delta(\epsilon))}.$$

The definitions and properties above can be generalized in the natural way to tuples of multiple random variables as well.

## 5.2 Random Codebook Generation

We recall that  $K$  denotes the collection  $(A, B, S)$ , and  $\mathcal{K}$  denotes the set  $\mathcal{A} \times \mathcal{B} \times \mathcal{S}$ .

Fix a distribution  $P_{UVABCSXYZS}$  from  $\mathcal{P}$  and a conditional distribution  $Q_{ABC|\tilde{U}\tilde{V}\tilde{K}\tilde{C}}$  satisfying (6), as required by the statement of the theorem. Fix a positive integer  $L$ . There are  $L$  blocks in encoding and decoding. Fix positive real numbers  $R'_1, R'_2, R_0, R_1, R_2, \rho_0, \rho_1$  and  $\rho_2$  such that  $R'_1 > R_1$  and  $R'_2 > R_2$ , where these numbers denote the rates of codebooks to be constructed as described below. Fix block length  $n$  and  $\epsilon > 0$ . Let  $\epsilon_l, l = 1, \dots, L$  be numbers such that  $\epsilon < \epsilon_1 < \epsilon_2 < \dots < \epsilon_L$ .

For  $l = 1, 2, 3, \dots, L$  independently perform the following random experiments.

- For each sequence  $\tilde{\mathbf{c}} \in \mathcal{C}^n$ , generate  $2^{n\rho_0}$  sequences  $\mathbf{C}_{[l,i,\tilde{\mathbf{c}}]}$ ,  $i = 1, 2, \dots, 2^{n\rho_0}$ , independently where each sequence is generated from the product distribution  $\prod_{i=1}^n P_{C|\tilde{C}}(\cdot|\tilde{c}_i)$ .
- For each sequence pair  $(\mathbf{c}, \mathbf{a}) \in \mathcal{C}^n \times \mathcal{A}^n$ , generate  $2^{n(R'_1 - R_1)}$  sequences  $\mathbf{U}_{[l,i,\mathbf{c},\mathbf{a}]}$ ,  $i = 1, 2, \dots, 2^{n(R'_1 - R_1)}$ , independently where each sequence is generated from the product distribution  $\prod_{i=1}^n P_{U|AC}(\cdot|a_i, c_i)$ . Call this the first  $U$ -bin. Independently repeat this experiment  $2^{nR_1}$  times to generate  $2^{nR_1}$   $U$ -bins, and a total of  $2^{nR'_1}$  sequences. The  $i$ th sequence in the  $j$ th bin is  $\mathbf{U}_{[l,(j-1)2^{nR_1}+i,\mathbf{c},\mathbf{a}]}$ .
- For each sequence pair  $(\mathbf{c}, \mathbf{b}) \in \mathcal{C}^n \times \mathcal{B}^n$ , similarly generate  $2^{nR_2}$   $V$ -bins each containing  $2^{n(R'_2 - R_2)}$  sequences with each sequence being generated from the product distribution  $\prod_{i=1}^n P_{V|BC}(\cdot|b_i, c_i)$ . The  $i$ th sequence in the  $j$ th bin is  $\mathbf{V}_{[l,(j-1)2^{nR_2}+i,\mathbf{c},\mathbf{b}]}$ .
- For each  $(\tilde{\mathbf{u}}, \tilde{\mathbf{c}}, \mathbf{c}) \in \mathcal{U}^n \times \mathcal{C}^n \times \mathcal{C}^n$  generate independently  $2^{n\rho_1}$  sequences  $\mathbf{A}_{[l,i,\tilde{\mathbf{u}},\tilde{\mathbf{c}},\mathbf{c}]}$ , for  $i = 1, 2, \dots, 2^{n\rho_1}$ , where each sequence is generated from  $\prod_{j=1}^n P_{A|\tilde{U}\tilde{C}C}(\cdot|\tilde{u}_j, \tilde{c}_j, c_j)$ .
- For each  $(\tilde{\mathbf{v}}, \tilde{\mathbf{c}}, \mathbf{c}) \in \mathcal{V}^n \times \mathcal{C}^n \times \mathcal{C}^n$  generate independently  $2^{n\rho_2}$  sequences  $\mathbf{B}_{[l,i,\tilde{\mathbf{v}},\tilde{\mathbf{c}},\mathbf{c}]}$ , for  $i = 1, 2, \dots, 2^{n\rho_2}$ , where each sequence is generated from  $\prod_{j=1}^n P_{B|\tilde{V}\tilde{C}C}(\cdot|\tilde{v}_j, \tilde{c}_j, c_j)$ .

- For each  $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}, \mathbf{v}) \in \mathcal{A}^n \times \mathcal{B}^n \times \mathcal{C}^n \times \mathcal{U}^n \times \mathcal{V}^n$  generate one sequence  $\mathbf{X}_{[l, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}, \mathbf{v}]}$  using  $\prod_{i=1}^n P_{X|ABCUV}(\cdot | a_i, b_i, c_i, u_i, v_i)$ .
- Generate independently sequences  $\mathbf{U}[0], \mathbf{V}[0], \mathbf{C}[0], \mathbf{K}[0], \mathbf{X}[0], \mathbf{Y}[0], \mathbf{Z}[0]$  from the product distribution  $P_{U,V,C,K,X,Y,Z}^n$ .

These sequences are known to all terminals before transmission begins.

### 5.3 Encoding Operation

Let  $W_0[l]$  denote the common message, and  $W_1[l], W_2[l]$ , the private messages for block  $l$ . These are independent random variables distributed uniformly over  $\{0, 1, \dots, 2^{nR_0} - 1\}$ ,  $\{1, 2, \dots, 2^{nR_1}\}$ , and  $\{1, 2, \dots, 2^{nR_2}\}$ , respectively. We set  $W_0[0] = W_1[0] = W_2[0] = W_0[L] = W_1[L] = W_2[L] = 1$ .

For each block  $l$ , the encoder chooses a quintuple of sequences  $(\mathbf{A}[l], \mathbf{B}[l], \mathbf{C}[l], \mathbf{U}[l], \mathbf{V}[l])$  from the five codebooks generated above, according to the encoding rule described below. The channel input, and channel output sequences in block  $l$  are denoted  $\mathbf{X}[l]$ ,  $\mathbf{Y}[l]$  and  $\mathbf{Z}[l]$ , respectively.

**Blocks  $l = 1, 2, 3, \dots, L$ :** The encoder performs the following sequence of operations.

- Step 1: The encoder determines a triplet of indices  $G_A[l] \in \{1, \dots, 2^{n\rho_1}\}$ ,  $G_B[l] \in \{1, \dots, 2^{n\rho_2}\}$ , and  $G_C[l] \in \{1, \dots, 2^{n\rho_0}\}$  such that

1.  $G_C[l] \bmod 2^{nR_0} = W_0[l]$ ,<sup>4</sup> and
2. The tuple  $(\mathbf{U}[l-1], \mathbf{V}[l-1], \mathbf{K}[l-1], \mathbf{C}[l-1])$  is jointly  $\epsilon_l$ -typical with the triplet of sequences

$$(\mathbf{C}_{[l, G_C[l], \mathbf{C}[l-1]]}, \mathbf{A}_{[l, G_A[l], \mathbf{U}[l-1], \mathbf{C}[l-1], \mathbf{C}_{[l, G_C[l], \mathbf{C}[l-1]]}], \mathbf{B}_{[l, G_B[l], \mathbf{V}[l-1], \mathbf{C}[l-1], \mathbf{C}_{[l, G_C[l], \mathbf{C}[l-1]]}]},$$

with respect to  $P_{\tilde{U}, \tilde{V}, \tilde{K}, \tilde{C}, C, A, B}$ .<sup>5</sup>

If no such index triplet is found, it declares error and sets  $(G_A[l], G_B[l], G_C[l]) = (1, 1, 1)$ .

The encoder then sets

$$\mathbf{C}[l] = \mathbf{C}_{[l, G_C[l], \mathbf{C}[l-1]]}, \mathbf{A}[l] = \mathbf{A}_{[l, G_A[l], \mathbf{U}[l-1], \mathbf{C}[l-1], \mathbf{C}_{[l, G_C[l], \mathbf{C}[l-1]]}], \mathbf{B}[l] = \mathbf{B}_{[l, G_B[l], \mathbf{V}[l-1], \mathbf{C}[l-1], \mathbf{C}_{[l, G_C[l], \mathbf{C}[l-1]]}]}$$

- Step 2: The encoder chooses a pair of indices  $(G_U[l], G_V[l])$  such that the triplet of sequences

$$(\mathbf{U}_{[l, G_U[l], \mathbf{C}[l], \mathbf{A}[l]]}, \mathbf{V}_{[l, G_V[l], \mathbf{C}[l], \mathbf{B}[l]]}, \mathbf{A}[l], \mathbf{B}[l], \mathbf{C}[l])$$

is  $\epsilon$ -typical with respect to  $P_{UVABC}$ , and  $\mathbf{U}_{[l, G_U[l], \mathbf{C}[l], \mathbf{A}[l]]}$  belongs to the  $U$ -bin with index  $W_1[l]$ , and  $\mathbf{V}_{[l, G_V[l], \mathbf{C}[l], \mathbf{B}[l]]}$  belongs to the  $V$ -bin with index  $W_2[l]$ . If no such index pair is found, it declares error and sets  $(G_U[l], G_V[l]) = (1, 1)$ .

The encoder then sets  $\mathbf{U}[l] = \mathbf{U}_{[l, G_U[l], \mathbf{C}[l], \mathbf{A}[l]]}$ ,  $\mathbf{V}[l] = \mathbf{V}_{[l, G_V[l], \mathbf{C}[l], \mathbf{B}[l]]}$ , and  $\mathbf{X}[l] = \mathbf{X}_{[l, \mathbf{A}[l], \mathbf{B}[l], \mathbf{C}[l], \mathbf{U}[l], \mathbf{V}[l]]}$ . It transmits  $\mathbf{X}[l]$  as the channel input sequence for block  $l$ .

- Step 3: The broadcast channel produces  $(\mathbf{Y}[l], \mathbf{Z}[l])$ .
- Step 4: After receiving  $(\mathbf{S}[l])$  via the feedback link, the encoder sets  $\mathbf{K}[l] = (\mathbf{A}[l], \mathbf{B}[l], \mathbf{S}[l])$ .

<sup>4</sup>This condition corresponds to the role of  $C$  in carrying the message  $W_0[l]$  common to both receivers.

<sup>5</sup> If there is more than one triplet satisfying the conditions, the encoder chooses one of them at random.

## 5.4 Decoding Operation

**Block 1:** The objective at the end of this block is to decode the common message  $W_0[1]$  at both receivers.

- The first decoder receives  $\mathbf{Y}[1]$ , and the second decoder receives  $\mathbf{Z}[1]$ .
- The first decoder determines the unique index pair  $(\hat{G}_{C1}[1], \hat{G}_A[1])$  such that the tuples

$$(\mathbf{C}[0], \mathbf{A}[0], \mathbf{U}[0], \mathbf{Y}[0]) \text{ and } (\bar{\mathbf{C}}_1[1], \mathbf{A}_{[1, \hat{G}_A[1], \mathbf{U}[0], \mathbf{C}[0], \bar{\mathbf{C}}_1[1]]}, \mathbf{Y}[1])$$

are jointly  $\epsilon_l$ -typical with respect to  $P_{\tilde{\mathbf{C}}\tilde{\mathbf{A}}\tilde{\mathbf{U}}\tilde{\mathbf{Y}}_{CAY}}$ , where  $\bar{\mathbf{C}}_1[1] \triangleq \mathbf{C}_{[1, \hat{G}_{C1}[1], \mathbf{C}[0]]}$ . Note that  $\bar{\mathbf{C}}_1[1]$  is the estimate of  $\mathbf{C}[1]$  at the first decoder.

If not successful in this operation, the first decoder declares an error and sets  $(\hat{G}_{C1}[1], \hat{G}_A[1]) = (1, 1)$ , and  $\bar{\mathbf{C}}_1[1] \triangleq \mathbf{C}_{[1, \hat{G}_{C1}[1], \mathbf{C}[0]]}$ .

- The first decoder outputs  $\hat{W}_0[1] = \hat{G}_{C1}[1] \bmod 2^{nR_0}$ , and sets

$$\bar{\mathbf{A}}[1] = \mathbf{A}_{[1, \hat{G}_A[1], \mathbf{U}[0], \mathbf{C}[0], \bar{\mathbf{C}}_1[1]]}.$$

$\bar{\mathbf{A}}[1]$  is the first decoder's estimate of  $\mathbf{A}[1]$ .

- The second decoder determines the unique index pair  $(\hat{G}_{C2}[1], \hat{G}_B[1])$  such that the tuples

$$(\mathbf{C}[0], \mathbf{B}[0], \mathbf{V}[0], \mathbf{Z}[0]) \text{ and } (\bar{\mathbf{C}}_2[1], \mathbf{B}_{[1, \hat{G}_B[1], \mathbf{V}[0], \mathbf{C}[0], \bar{\mathbf{C}}_2[1]]}, \mathbf{Z}[1])$$

are jointly  $\epsilon_l$ -typical with respect to  $P_{\tilde{\mathbf{C}}\tilde{\mathbf{B}}\tilde{\mathbf{V}}\tilde{\mathbf{Z}}_{CBZ}}$ , where  $\bar{\mathbf{C}}_2[1] \triangleq \mathbf{C}_{[1, \hat{G}_{C2}[1], \mathbf{C}[0]]}$ . Note that  $\bar{\mathbf{C}}_2[1]$  is the estimate of  $\mathbf{C}[1]$ , at the second decoder.

If not successful in this operation, the second decoder declares an error and sets  $(\hat{G}_{C2}[1], \hat{G}_B[1]) = (1, 1)$ , and  $\bar{\mathbf{C}}_2[1] \triangleq \mathbf{C}_{[1, \hat{G}_{C2}[1], \mathbf{C}[0]]}$ .

- The second decoder outputs  $\bar{W}_0[1] = \hat{G}_{C2}[1] \bmod 2^{nR_0}$ , and sets

$$\bar{\mathbf{B}}[1] = \mathbf{B}_{[1, \hat{G}_B[1], \mathbf{V}[0], \mathbf{C}[0], \bar{\mathbf{C}}_2[1]]}.$$

$\bar{\mathbf{B}}[1]$  is the second decoder's estimate of  $\mathbf{B}[1]$ .

**Block  $l, l = 2, 3, \dots, L$ :** The objective at the end of block  $l$  is for receiver 1 to decode  $(W_0[l], W_1[l-1])$  and for receiver 2 to decode  $(W_0[l], W_2[l-1])$ .

- The first decoder receives  $\mathbf{Y}[l]$  and the second decoder receives  $\mathbf{Z}[l]$ .
- The first decoder determines the unique index triplet  $(\hat{G}_{C1}[l], \hat{G}_A[l], \hat{G}_U[l-1])$  such that the tuples

$$(\bar{\mathbf{C}}_1[l-1], \bar{\mathbf{A}}[l-1], \bar{\mathbf{U}}[l-1], \mathbf{Y}[l-1]) \text{ and } (\bar{\mathbf{C}}_1[l], \mathbf{A}_{[l, \hat{G}_A[l], \bar{\mathbf{U}}[l-1], \bar{\mathbf{C}}_1[l-1], \bar{\mathbf{C}}_1[l]], \mathbf{Y}[l])$$

are jointly  $\epsilon_l$ -typical with respect to  $P_{\tilde{\mathbf{C}}\tilde{\mathbf{A}}\tilde{\mathbf{U}}\tilde{\mathbf{Y}}_{CAY}}$ , where

$$\bar{\mathbf{U}}[l-1] \triangleq \mathbf{U}_{[(l-1), \hat{G}_U[l-1], \bar{\mathbf{C}}_1[l-1], \bar{\mathbf{A}}[l-1]]}, \quad \bar{\mathbf{C}}_1[l] \triangleq \mathbf{C}_{[l, \hat{G}_{C1}[l], \bar{\mathbf{C}}_1[l-1]]}.$$

If not successful in this operation, the first decoder declares an error and sets  $(\hat{G}_{C1}[l], \hat{G}_A[l], \hat{G}_U[l-1]) = (1, 1, 1)$ , and

$$\bar{\mathbf{U}}[l-1] = \mathbf{U}_{[(l-1), 1, \bar{\mathbf{C}}_1[l-1], \bar{\mathbf{A}}[l-1]]}, \quad \bar{\mathbf{C}}_1[l] \triangleq \mathbf{C}_{[l, 1, \bar{\mathbf{C}}_1[l-1]]}.$$

Note that  $\bar{\mathbf{U}}[l-1]$  and  $\bar{\mathbf{C}}_1[l]$  are the estimates of  $\mathbf{U}[l-1]$  and  $\mathbf{C}[l]$ , respectively, at the first decoder.

- The first decoder then outputs  $\hat{W}_0[l] = \hat{G}_{C1}[l] \bmod 2^{nR_0}$ , and  $\hat{W}_1[l-1]$  as the index of  $U$ -bin that contains the sequence  $\mathbf{U}_{[(l-1), \hat{G}_U[l-1], \bar{\mathbf{C}}_1[l-1], \bar{\mathbf{A}}[l-1]]}$ . The decoder sets

$$\bar{\mathbf{A}}[l] = \mathbf{A}_{[l, \hat{G}_A[l], \bar{\mathbf{U}}[l-1], \bar{\mathbf{C}}_1[l-1], \bar{\mathbf{C}}_1[l]]}.$$

$\bar{\mathbf{A}}[l]$  is the first decoder's estimate of  $\mathbf{A}[l]$ .

- The second decoder determines the unique index triplet  $(\hat{G}_{C2}[l], \hat{G}_B[l], \hat{G}_V[l-1])$  such that the tuples

$$(\bar{\mathbf{C}}_2[l-1], \bar{\mathbf{B}}[l-1], \bar{\mathbf{V}}[l-1], \mathbf{Z}[l-1]) \quad \text{and} \quad (\bar{\mathbf{C}}_2[l], \mathbf{B}_{[l, \hat{G}_B[l], \bar{\mathbf{V}}[l-1], \bar{\mathbf{C}}_2[l-1], \bar{\mathbf{C}}_2[l]]}, \mathbf{Z}[l])$$

are jointly  $\epsilon_l$ -typical with respect to  $P_{\bar{\mathbf{C}}\bar{\mathbf{B}}\bar{\mathbf{V}}\bar{\mathbf{Z}}\mathbf{C}\mathbf{B}\mathbf{Z}}$ , where, where

$$\bar{\mathbf{V}}[l-1] \triangleq \mathbf{V}_{[(l-1), \hat{G}_V[l-1], \bar{\mathbf{C}}_2[l-1], \bar{\mathbf{B}}[l-1]]}, \quad \bar{\mathbf{C}}_2[l] \triangleq \mathbf{C}_{[l, \hat{G}_{C2}[l], \bar{\mathbf{C}}_2[l-1]]}.$$

If not successful in this operation, the second decoder declares an error and sets  $(\hat{G}_{C2}[l], \hat{G}_B[l], \hat{G}_V[l-1]) = (1, 1, 1)$ , and

$$\bar{\mathbf{V}}[l-1] \triangleq \mathbf{V}_{[(l-1), 1, \bar{\mathbf{C}}_2[l-1], \bar{\mathbf{B}}[l-1]]}, \quad \bar{\mathbf{C}}_2[l] \triangleq \mathbf{C}_{[l, 1, \bar{\mathbf{C}}_2[l-1]]};$$

Note that  $\bar{\mathbf{V}}[l-1]$  and  $\bar{\mathbf{C}}_2[l]$  are the estimates of  $\mathbf{V}[l-1]$  and  $\mathbf{C}[l]$ , respectively, at the second decoder.

- The second decoder then outputs  $\bar{W}_0[l] = \hat{G}_{C2}[l] \bmod 2^{nR_0}$ , and  $\bar{W}_2[l-1]$  as the index of  $V$ -bin that contains the sequence  $\mathbf{V}_{[(l-1), \hat{G}_V[l-1], \bar{\mathbf{C}}_2[l-1], \bar{\mathbf{B}}[l-1]]}$ . The decoder sets

$$\bar{\mathbf{B}}[l] = \mathbf{B}_{[l, \hat{G}_B[l], \bar{\mathbf{V}}[l-1], \bar{\mathbf{C}}_2[l-1], \bar{\mathbf{C}}_2[l]]}.$$

$\bar{\mathbf{B}}[l]$  is the second decoder's estimate of  $\mathbf{B}[l]$ .

## 5.5 Error Analysis

Let  $\mathcal{E}[0]$  denote the event that  $(\mathbf{U}[0], \mathbf{K}[0], \mathbf{V}[0], \mathbf{C}[0])$  is not  $\epsilon[0]$ -typical with respect to  $P_{UKVC}$ . By Property 0, we have  $\Pr[\mathcal{E}[0]] \leq \epsilon$  for all sufficiently large  $n$ .

**Block 1:** The error event in Block 1 can be expressed as  $\mathcal{E}[1] = \mathcal{E}_1[1] \cup \mathcal{E}_2[1] \cup \mathcal{E}_3[1] \cup \mathcal{E}_4[1] \cup \mathcal{E}_5[1]$  where

- $\mathcal{E}_1[1]$  is the event that the encoder declares error in step 1 of encoding (described in Section 5.3),
- $\mathcal{E}_2[1]$  is the event that the encoder declares error in step 2 of encoding,
- $\mathcal{E}_3[1]$  is the event that the tuples  $(\mathbf{U}[0], \mathbf{V}[0], \mathbf{K}[0], \mathbf{C}[0])$  and  $(\mathbf{U}[1], \mathbf{V}[1], \mathbf{K}[1], \mathbf{C}[1])$  are not jointly  $\epsilon_1$ -typical with respect to  $P_{\bar{\mathbf{U}}\bar{\mathbf{V}}\bar{\mathbf{K}}\bar{\mathbf{C}}\mathbf{U}\mathbf{V}\mathbf{K}\mathbf{C}}$ ,
- $\mathcal{E}_4[1]$  is the event that  $(\hat{G}_{C1}[1], \hat{G}_A[1]) \neq (G_C[1], G_A[1])$ , and  $\mathcal{E}_5[1]$  is the event that  $(\hat{G}_{C2}[1], \hat{G}_B[1]) \neq (G_C[1], G_B[1])$ .

**Lemma 5.1** (Covering lemma).  $\Pr[\mathcal{E}_1[1] \mid \mathcal{E}[0]^c] \leq \epsilon$  for all sufficiently large  $n$  if  $R_0, \rho_0, \rho_1$ , and  $\rho_2$  satisfy

$$\rho_0 > I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + \delta(\epsilon_1) \quad (50)$$

$$\rho_0 + \rho_1 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + \delta(\epsilon_1) \quad (51)$$

$$\rho_0 + \rho_2 > I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + \delta(\epsilon_1) \quad (52)$$

$$\rho_0 + \rho_1 + \rho_2 > I(\tilde{V}\tilde{K}; A|C\tilde{C}\tilde{U}) + I(\tilde{U}\tilde{K}; B|C\tilde{C}\tilde{V}) + I(A; B|\tilde{U}\tilde{V}\tilde{K}C\tilde{C}) + I(\tilde{U}\tilde{V}\tilde{K}; C|\tilde{C}) + R_0 + \delta(\epsilon_1) \quad (53)$$

*Proof.* The proof of this covering lemma is the same as that of [17, Lemma 14.1], with  $(\tilde{U}, \tilde{K})$  and  $(\tilde{V}, \tilde{K})$  assuming the roles of the two sources being covered.  $\square$

**Lemma 5.2.**  $\Pr[\mathcal{E}_2[1] \mid \mathcal{E}[0]^c] \leq \epsilon$  for all sufficiently large  $n$  if  $R'_1, R'_2$ , and  $R_1, R_2$  satisfy

$$R'_1 + R'_2 - R_1 - R_2 > H(U|AC) + H(V|BC) - H(UV|ABC) + \delta(\epsilon_1) \quad (54)$$

*Proof.* This is very similar to a standard covering lemma used for bounding the probability of encoding error in Marton's coding scheme, a proof of which can be found in [2, 16] or [17].  $\square$

From Property 1 of typical sequences, it follows that  $\Pr[\mathcal{E}_3[1] \mid \mathcal{E}_1[1]^c, \mathcal{E}_2[1]^c, \mathcal{E}[0]^c] \leq \epsilon$  for all sufficiently large  $n$ .

**Lemma 5.3.**  $\Pr[\mathcal{E}_4[1] \cup \mathcal{E}_5[1] \mid \mathcal{E}_3[1]^c, \mathcal{E}_2[1]^c, \mathcal{E}_1[1]^c, \mathcal{E}[0]^c] \leq 2\epsilon$ , if

$$\rho_0 + \rho_1 < I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - 3\delta(\epsilon_1) \quad (55)$$

$$\rho_0 + \rho_2 < I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - 3\delta(\epsilon_1) \quad (56)$$

$$\rho_1 < I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - 3\delta(\epsilon_1) \quad (57)$$

$$\rho_2 < I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - 3\delta(\epsilon_1) \quad (58)$$

*Proof.* The proof is a special case of that of Lemma 5.4 given below.  $\square$

Hence  $P[\mathcal{E}[1] \mid \mathcal{E}[0]^c] < 5\epsilon$  if the conditions given in Lemmas 5.1, 5.2, and 5.3 are satisfied. This implies that  $\bar{\mathbf{A}}[1] = \mathbf{A}[1]$ ,  $\bar{\mathbf{C}}_1[1] = \bar{\mathbf{C}}_2[1] = \mathbf{C}[1]$ , and similarly  $\bar{\mathbf{B}}[1] = \mathbf{B}[1]$  with high probability.

**Block  $l$ ,  $l = 2, 3, \dots, L$ :** The error event in block  $l$  can be expressed as  $\mathcal{E}[l] = \cup_{i=1}^5 \mathcal{E}_i[l]$  where

- $\mathcal{E}_1[l]$  is the event that the encoder declares error in step 1 of encoding
- $\mathcal{E}_2[l]$  is the event that the encoder declares error in step 2 of encoding,
- $\mathcal{E}_3[l]$  is the event that the tuples  $(\mathbf{U}[l-1], \mathbf{V}[l-1], \mathbf{K}[l-1], \mathbf{C}[l-1])$  and  $(\mathbf{U}[l], \mathbf{V}[l], \mathbf{K}[l], \mathbf{C}[l])$  are not jointly  $\epsilon_l$ -typical with respect to  $P_{\tilde{U}\tilde{V}\tilde{K}\tilde{C}UVKC}$ ,
- $\mathcal{E}_4[l]$  is the event that  $\{(\hat{G}_{C1}[l], \hat{G}_A[l], \hat{G}_U[l-1]) \neq (G_C[l], G_A[l], G_U[l-1])\}$ , and  $\mathcal{E}_5[l]$  is the event that  $\{(\hat{G}_{C2}[l], \hat{G}_B[l], \hat{G}_V[l-1]) \neq (G_C[l], G_B[l], G_V[l-1])\}$ .

Using arguments similar to those used for Block 1, one can show that if  $\rho_0, \rho_1, \rho_2, R'_1, R'_2, R_1$  and  $R_2$  satisfy the conditions given in (54) and (50)-(53) with  $\epsilon[l-1]$  replaced with  $\epsilon_l$ , then for all sufficiently large  $n$ ,

$$\Pr[\mathcal{E}_1[l] \cup \mathcal{E}_2[l] \cup \mathcal{E}_3[l] \mid \cap_{k=0}^{l-1} \mathcal{E}[k]^c] \leq 3\epsilon.$$

**Lemma 5.4** (Packing lemma).  *$\Pr[\mathcal{E}_4[l] \cup \mathcal{E}_5[l] \mid \mathcal{E}_3[l]^c, \mathcal{E}_2[l]^c, \mathcal{E}_1[l]^c, \cap_{k=0}^{l-1} \mathcal{E}[k]^c] \leq 2\epsilon$ , if*

$$R'_1 + \rho_0 + \rho_1 < I(\tilde{U}; Y\tilde{Y}\tilde{A}|\tilde{C}) + I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - 3\delta(\epsilon_l) \quad (59)$$

$$R'_1 + \rho_1 < I(\tilde{U}; Y\tilde{A}\tilde{Y}C|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - 3\delta(\epsilon_l) \quad (60)$$

$$R'_2 + \rho_0 + \rho_2 < I(\tilde{V}; Z\tilde{Z}\tilde{B}|\tilde{C}) + I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - 3\delta(\epsilon_l) \quad (61)$$

$$R'_2 + \rho_2 < I(\tilde{V}; Z\tilde{B}\tilde{Z}C|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - 3\delta(\epsilon_l) \quad (62)$$

$$\rho_0 + \rho_1 < I(C; Y\tilde{A}\tilde{Y}\tilde{U}|\tilde{C}) + I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - 3\delta(\epsilon_l) \quad (63)$$

$$\rho_0 + \rho_2 < I(C; Z\tilde{B}\tilde{Z}\tilde{V}|\tilde{C}) + I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - 3\delta(\epsilon_l) \quad (64)$$

$$\rho_1 < I(A; Y\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - 3\delta(\epsilon_l) \quad (65)$$

$$\rho_2 < I(B; Z\tilde{B}\tilde{Z}|\tilde{V}\tilde{C}C) - 3\delta(\epsilon_l) \quad (66)$$

*Proof.* See Appendix C. □

Hence  $\Pr[\mathcal{E}[l] \mid \cap_{k=0}^{l-1} \mathcal{E}[k]^c] < 5\epsilon$ . Under the event  $\cap_{k=0}^l \mathcal{E}[k]^c$ , we have  $\bar{\mathbf{A}}[l] = \mathbf{A}[l]$ ,  $\bar{\mathbf{C}}_1[l] = \bar{\mathbf{C}}_2[l] = \mathbf{C}[l]$ , and  $\bar{\mathbf{B}}[l] = \mathbf{B}[l]$ .

**Overall Probability of Decoding Error** : Hence the probability of decoding error over  $L$  blocks satisfies

$$\Pr[\mathcal{E}] = \Pr[\cup_{l=0}^L \mathcal{E}[l]] \leq 5\epsilon L$$

if the conditions given in (54), (50)-(53) and (59)-(66) are satisfied with  $\delta(\epsilon_1)$  and  $\delta(\epsilon_l)$  are replaced with  $\theta$ , where  $\theta = \sum_{l=1}^L \delta(\epsilon_l)$ . This implies that the rate region given by (14), (15)-(18), (19)-(26) is achievable. By applying Fourier-Motzkin elimination to these equations, we obtain that the rate region given in the statement of the theorem is achievable. The details of this elimination are omitted since they are elementary, but somewhat tedious.

## 6 Conclusion

We have derived a single-letter rate region for the two-user broadcast channel with feedback. Using the Marton coding scheme as the starting point, our scheme has a block-Markov structure and uses three additional random variables  $(A, B, C)$  to cover the correlated information generated at the end of each block. The proposed region was used to compute achievable rates for the AWGN channel with noisy feedback. In particular, it was shown that sum rates higher than the no-feedback sum capacity could be achieved even with noisy feedback to only one receiver.

The key to obtaining a single-letter characterization was to impose a constraint on the Markov kernel connecting the distribution of the random variables across successive blocks. A similar idea was used in [19]



for multiple-access channels with feedback. This approach to harnessing correlated information is quite general, and it is likely that it can be used to obtain improved rate regions for other multi-user channels with feedback such as interference and relay channels.

## Acknowledgements

We thank the associate editor and the anonymous reviewers for their comments and suggestions, which led to a much improved the paper.

## References

- [1] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 306–311, 1979.
- [2] T. M. Cover, “Comments on broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, 1998.
- [3] A. El Gamal, “The feedback capacity of degraded broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 24, pp. 379–381, May 1980.
- [4] G. Dueck, “Partial feedback for two-way and broadcast channels,” *Inform. and Control*, vol. 46, pp. 1–15, July 1980.
- [5] L. H. Ozarow and S. K. Leung-Yan-Cheong, “An achievable region and outer bound for the Gaussian broadcast channel with feedback,” *IEEE Trans. on Inform. Theory*, vol. 30, no. 4, pp. 667–671, 1984.
- [6] S. R. Bhaskaran, “Gaussian broadcast channel with feedback,” *IEEE Trans. on Inform. Theory*, vol. 54, no. 11, pp. 5252–5257, 2008.
- [7] G. Kramer, “Capacity Results for the Discrete Memoryless Network,” *IEEE Trans Inf Theory*, vol. 49, pp. 4–20, January 2003.
- [8] T. S. Han and M. H. M. Costa, “Broadcast channels with arbitrarily correlated sources,” *IEEE Trans. Inform. Theory*, vol. IT-33, no. 5, pp. 641–650, 1987.
- [9] G. Kramer and C. Nair, “Comments on ‘Broadcast channels with arbitrarily correlated sources’,” in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2009.
- [10] P. Minero and Y.-H. Kim, “Correlated sources over broadcast channels,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2009.
- [11] S. Choi and S. S. Pradhan, “A graph-based framework for transmission of correlated sources over broadcast channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2841–2856, 2008.
- [12] W. Kang and G. Kramer, “Broadcast channel with degraded source random variables and receiver side information,” in *Proc. IEEE Int. Symp. on Inf. Theory*, June 2008.

- [13] M. Gastpar, A. Lapidoth, Y. Steinberg, and M. A. Wigger, “Feedback can double the prelog of some memoryless gaussian networks,” *submitted to IEEE Trans Inf Theory*, 2010. <http://arxiv.org/abs/1003.6082>.
- [14] R. Venkataramanan and S. S. Pradhan, “Achievable rates for the broadcast channel with feedback,” *Proc. IEEE Int. Symp. on Inf. Theory*, 2010.
- [15] O. Shayevitz and M. A. Wigger, “On the capacity of the discrete memoryless broadcast channel with feedback,” *Proc. IEEE Int. Symp. on Inf. Theory*, 2010. <http://arxiv.org/abs/1012.6012>.
- [16] A. El Gamal and E. C. van der Muelen, “A proof of Marton’s coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, 1981.
- [17] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.
- [18] T. S. Han, “A general coding scheme for the two-way channel,” *IEEE Trans. Inform. Theory*, vol. 30, no. 1, pp. 35–43, 1984.
- [19] R. Venkataramanan and S. S. Pradhan, “A new achievable rate region for the multiple-access channel with noiseless feedback,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 8038–8054, Dec. 2011.
- [20] J. P. M. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedback- part II: Band-limited signals,” *IEEE Trans. on Information Theory*, vol. IT-12, pp. 183–189, April 1966.
- [21] Y.-H. Kim, A. Lapidoth, and T. Weissman, “The gaussian channel with noisy feedback,” in *IEEE Int. Symp. on Information Theory*, June 2007.
- [22] T. Cover, “Broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 18, pp. 2 – 14, Jan 1972.
- [23] P. Bergmans, “A simple converse for broadcast channels with additive white gaussian noise (corresp.),” *IEEE Trans. Inform. Theory*, vol. 20, pp. 279 – 280, Mar 1974.
- [24] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. 29, pp. 439 – 441, May 1983.
- [25] S. I. Gelfand and M. S. Pinsker, “Coding for channel with random parameters,” *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19 – 31, 1980.
- [26] R. M. Gray and A. D. Wyner, “Source coding for a simple network,” *Bell System Technical Journal*, vol. 53, pp. 1681 – 1721, Nov. 1974.
- [27] A. Orlitsky and J. Roche, “Coding for computing,” *IEEE Trans. Inf. Theory*, vol. 47, pp. 903 – 917, Mar 2001.
- [28] S. H. Lim, P. Minero, and Y.-H. Kim, “Lossy communication of correlated sources over multiple access channels,” in *48th Allerton Conf. on Communication, Control, and Computing*, pp. 851 –858, 2010.

## APPENDIX

### A Mutual Information terms for the AWGN example

With the joint distribution described in Section 4.2, we first compute the following quantities.

$$M_u \triangleq E[\Delta^2] = \alpha P_1 \left( \frac{\sigma^2 + \bar{\beta} \bar{\alpha} P_1}{P_1 + \sigma^2} \right)^2 + \bar{\alpha} P_1 \left( \frac{\beta \sigma^2 - \alpha \bar{\beta} P_1}{P_1 + \sigma^2} \right)^2 + \left( \frac{\alpha P_1 + \beta \bar{\alpha} P_1}{P_1 + \sigma^2} \right)^2 \frac{\sigma^4}{\sigma^2 + \sigma_f^2}, \quad (67)$$

$$E[\Delta \tilde{V}] = \frac{\bar{\alpha} P_1 (\beta \sigma^2 - \alpha \bar{\beta} P_1)}{P_1 + \sigma^2} \quad (68)$$

$$E[\Delta \tilde{U}] = \frac{\alpha P_1 (\sigma^2 + \bar{\alpha} \bar{\beta} P_1)}{P_1 + \sigma^2} + \frac{\beta \bar{\alpha} P_1 (\beta \sigma^2 - \alpha \bar{\beta} P_1)}{P_1 + \sigma^2} \quad (69)$$

$$E[\Delta \tilde{Z}] = \frac{P_1 \sigma^2 (\alpha + \bar{\alpha} \beta)}{P_1 + \sigma^2} \left( 1 - \frac{\sigma^2 \rho}{\sigma^2 + \sigma_f^2} \right) \quad (70)$$

$$E[\Delta \tilde{Y}] = \frac{P_1 \sigma^2 (\alpha + \bar{\alpha} \beta)}{P_1 + \sigma^2} \left( \frac{\sigma_f^2}{\sigma^2 + \sigma_f^2} \right). \quad (71)$$

We next compute the conditional variances in terms of which the mutual information quantities are expressed.

$$\text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{V}) = 1 - \frac{(E[\Delta \tilde{V}])^2}{M_u \bar{\alpha} P_1} \quad (72)$$

$$\text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{U}) = 1 - \frac{(E[\Delta \tilde{U}])^2}{M_u (\alpha P_1 + \beta^2 \bar{\alpha} P_1)} \quad (73)$$

$$\text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{Z}) = 1 - \frac{(E[\Delta \tilde{Z}])^2}{M_u (P_1 + \sigma^2)} \quad (74)$$

$$\text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{Y}) = 1 - \frac{(E[\Delta \tilde{Y}])^2}{M_u (P_1 + \sigma^2)} \quad (75)$$

$$\text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{V} \tilde{Z}) = 1 - \frac{1}{M_u} \left( \frac{a_1 E[\Delta \tilde{V}] + b_1 E[\Delta \tilde{Z}]}{\bar{\alpha} P_1 (\alpha P_1 + \sigma^2)} \right) \quad (76)$$

$$\text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{U} \tilde{Y}) = 1 - \frac{1}{M_u} \left( \frac{a_2 E[\Delta \tilde{U}] + b_2 E[\Delta \tilde{Y}]}{(P_1 + \sigma^2)(\alpha P_1 + \beta^2 \bar{\alpha} P_1) - (\alpha P_1 + \beta \bar{\alpha} P_1)^2} \right) \quad (77)$$

$$(78)$$

where

$$\begin{aligned} a_1 &= E[\Delta \tilde{V}](P_1 + \sigma^2) - E[\Delta \tilde{Z}] \bar{\alpha} P_1, & b_1 &= E[\Delta \tilde{Z}] \bar{\alpha} P_1 - E[\Delta \tilde{V}] \bar{\alpha} P_1, \\ a_2 &= E[\Delta \tilde{U}](P_1 + \sigma^2) - E[\Delta \tilde{Y}] (\alpha P_1 + \beta \bar{\alpha} P_1), & b_2 &= E[\Delta \tilde{Y}] (\alpha P_1 + \beta^2 \bar{\alpha} P_1) - E[\Delta \tilde{U}] (\alpha P_1 + \beta \bar{\alpha} P_1). \end{aligned} \quad (79)$$

Finally, the mutual information terms are calculated to be

$$\begin{aligned}
I(U; V) &= \frac{1}{2} \log_2 \left( 1 + \frac{\beta^2 \bar{\alpha}}{\alpha} \right), \\
I(U; Y|C_0) &= \frac{1}{2} \log_2 \left( \frac{(P_1 + \sigma^2)(\alpha + \beta^2 \bar{\alpha})}{(P_1 + \sigma^2)(\alpha + \beta^2 \bar{\alpha}) - P_1(\alpha + \beta \bar{\alpha})^2} \right), \quad I(V; Z|C_0) = \frac{1}{2} \log_2 \left( \frac{P_1 + \sigma^2}{\alpha P_1 + \sigma^2} \right), \\
I(C_0; \tilde{U} \tilde{S} | \tilde{C}_0 \tilde{V}) &= \frac{1}{2} \log_2 \left( 1 + \frac{(1-D)}{D} \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{V}) \right), \quad I(C_0; \tilde{V} \tilde{S} | \tilde{C}_0 \tilde{U}) = \frac{1}{2} \log_2 \left( 1 + \frac{(1-D)}{D} \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{U}) \right), \\
I(C_0; \tilde{Y} | \tilde{C}_0 \tilde{U}) &= \frac{1}{2} \log_2 \left( \frac{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{U}) + D}{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{U} \tilde{Y}) + D} \right), \quad I(C_0; \tilde{Z} | \tilde{C}_0 \tilde{V}) = \frac{1}{2} \log_2 \left( \frac{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{V}) + D}{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{V} \tilde{Z}) + D} \right), \\
I(C_0; \tilde{U} | \tilde{C}_0 \tilde{Y}) &= \frac{1}{2} \log_2 \left( \frac{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{Y}) + D}{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{U} \tilde{Y}) + D} \right), \quad I(C_0; \tilde{V} | \tilde{C}_0 \tilde{Z}) = \frac{1}{2} \log_2 \left( \frac{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{Z}) + D}{(1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{V} \tilde{Z}) + D} \right), \\
I(C_0; Y | \tilde{C}_0 \tilde{Y}) &= \frac{1}{2} \log_2 \left( 1 + \frac{(P - P_1)((1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{Y}) + D)}{P_1 + \sigma^2} \right), \\
I(C_0; Z | \tilde{C}_0 \tilde{Z}) &= \frac{1}{2} \log_2 \left( 1 + \frac{(P - P_1)((1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{Z}) + D)}{P_1 + \sigma^2} \right), \\
I(C_0; Y | \tilde{C}_0 \tilde{U} \tilde{Y}) &= \frac{1}{2} \log_2 \left( 1 + \frac{(P - P_1)((1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{U} \tilde{Y}) + D)}{P_1 + \sigma^2} \right), \\
I(C_0; Z | \tilde{C}_0 \tilde{V} \tilde{Z}) &= \frac{1}{2} \log_2 \left( 1 + \frac{(P - P_1)((1-D) \text{var}(\tilde{T}_1 | \tilde{C}_0 \tilde{V} \tilde{Z}) + D)}{P_1 + \sigma^2} \right).
\end{aligned}$$

## B Comparison with Shayevitz-Wigger Region

In Theorem 1, set  $A = \tilde{V}_1, B = \tilde{V}_2, C = (\tilde{V}_0, W)$  and consider joint distributions over two blocks of the form

$$P_{\tilde{U} \tilde{V} \tilde{W} \tilde{X} \tilde{Y} \tilde{Z} \tilde{S}} Q_{\tilde{V}_0 \tilde{V}_1 \tilde{V}_2 | \tilde{U} \tilde{V} \tilde{W} \tilde{S}} P_{WUV} P_{X|WUV} P_{YZS|X}. \quad (80)$$

If we set  $Q_{\tilde{V}_0 \tilde{V}_1 \tilde{V}_2 | \tilde{U} \tilde{V} \tilde{W} \tilde{S}} = P_{V_0|UVWS} P_{V_1|V_0UVWS} P_{V_2|V_0UVWS}$ , the joint distribution is identical to that of the Shayevitz-Wigger region. With this distribution, Theorem 1 yields the following. (The parts in bold indicate the corresponding constraints of the Shayevitz-Wigger region.)

$$\begin{aligned}
R_0 &\leq \{\mathcal{T}_1, \mathcal{T}_2\} \\
R_0 + R_1 &\leq \mathbf{I}(\mathbf{UW}; \mathbf{YV}_1) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_0 \mathbf{V}_1 | \mathbf{Y}) + I(V_0; UW | V_1 Y) \\
R_0 + R_2 &\leq \mathbf{I}(\mathbf{VW}; \mathbf{ZV}_2) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_0 \mathbf{V}_2 | \mathbf{Y}) + I(V_0; VW | V_2 Z) \\
R_0 + R_1 + R_2 &\leq \mathbf{I}(\mathbf{UW}; \mathbf{YV}_1) + \mathbf{I}(\mathbf{V}; \mathbf{ZV}_2 | \mathbf{W}) - \mathbf{I}(\mathbf{U}; \mathbf{V} | \mathbf{W}) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_0 \mathbf{V}_1 | \mathbf{Y}) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_2 | \mathbf{V}_0 \mathbf{Z}) \\
&\quad + I(V_0; UW | V_1 Y) + I(V_0; V | V_2 W Z) + I(V_2; W | Z V_0) \\
R_0 + R_1 + R_2 &\leq \mathbf{I}(\mathbf{VW}; \mathbf{ZV}_2) + \mathbf{I}(\mathbf{U}; \mathbf{YV}_1 | \mathbf{W}) - \mathbf{I}(\mathbf{U}; \mathbf{V} | \mathbf{W}) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_0 \mathbf{V}_2 | \mathbf{Z}) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_1 | \mathbf{V}_0 \mathbf{Y}) \\
&\quad + I(V_0; VW | V_2 Z) + I(V_0; U | V_1 W Y) + I(V_1; W | Y V_0) \\
2R_0 + R_1 + R_2 &\leq \mathbf{I}(\mathbf{UW}; \mathbf{YV}_1) + \mathbf{I}(\mathbf{VW}; \mathbf{ZV}_2) - \mathbf{I}(\mathbf{U}; \mathbf{V} | \mathbf{W}) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_0 \mathbf{V}_1 | \mathbf{Y}) - \mathbf{I}(\mathbf{UVWS}; \mathbf{V}_0 \mathbf{V}_2 | \mathbf{Z}) \\
&\quad + I(V_0; UW | V_1 Y) + I(V_0; VW | V_2 Z)
\end{aligned} \quad (81)$$

where

$$\begin{aligned}\mathcal{T}_1 &= I(W; Y) + I(V_1 V_0; Y|WU) - I(VS; V_1 V_0|WU) + I(V_2; Z|WV V_0) - I(US; V_2|WV V_0) \\ \mathcal{T}_2 &= I(W; Z) + I(V_2 V_0; Z|WV) - I(US; V_2 V_0|WV) + I(V_1; Y|WU V_0) - I(VS; V_1|WU V_0)\end{aligned}\quad (82)$$

## C Proof of Lemma 5.4

We show through induction that if  $\Pr[\mathcal{E}_4[k]] < \epsilon$  for  $k = 1, \dots, l-1$ , then  $\Pr(\mathcal{E}_4[l]) < \epsilon$  if the conditions in the statement of the lemma are satisfied.

For conciseness, let  $\mathcal{F}$  denote the event  $(\cap_{k=0}^{l-1} \mathcal{E}[k]^c \cap \mathcal{E}_1[l]^c \cap \mathcal{E}_2[l]^c \cap \mathcal{E}_3[l]^c)$ . Note that  $\mathcal{F}$  is the conditioning event in the statement of Lemma 5.4; hence  $\bar{\mathbf{C}}_1[l-1] = \bar{\mathbf{C}}_2[l-1] = \mathbf{C}[l-1]$ ,  $\bar{\mathbf{A}}[l-1] = \mathbf{A}[l-1]$  and  $\bar{\mathbf{B}}[l-1] = \mathbf{B}[l-1]$ . Recall that given  $\mathbf{C}[l-1]$  and the indices  $G_C[l], G_A[l], G_U[l-1]$ , the following sequences are determined:

$$\begin{aligned}\mathbf{U}[l-1] &= \mathbf{U}_{[l-1, G_U[l-1], \mathbf{C}[l-1], \mathbf{A}[l-1]]}, \\ \mathbf{C}[l] &= \mathbf{C}_{[l, G_C[l], \mathbf{C}[l-1]]}, \\ \mathbf{A}[l] &= \mathbf{A}_{[l, G_A[l], \mathbf{U}[l-1], \mathbf{C}[l-1], \mathbf{C}[l]]}.\end{aligned}\quad (83)$$

Define the following indicator random variable:  $\psi(i, j, k) = 1$  if the tuples

$$(\mathbf{U}_{[l-1, k, \mathbf{C}[l-1]], \mathbf{A}[l-1]], \mathbf{A}[l-1], \mathbf{Y}[l-1], \mathbf{C}[l-1]) \quad \text{and} \quad (\mathbf{C}_{[l, i, \mathbf{C}[l-1]]}, \mathbf{Y}[l], \mathbf{A}_{[l, j, \mathbf{U}_{[l-1, k, \mathbf{C}[l-1]], \mathbf{A}[l-1]], \mathbf{C}[l-1], \mathbf{C}_{[l, i, \mathbf{C}[l-1]]}]})$$

are jointly  $\epsilon_l$ -typical with respect to  $P_{\tilde{U}\tilde{A}\tilde{Y}\tilde{C}CA\tilde{Y}}$  and 0 otherwise. We have

$$\begin{aligned}\Pr(\mathcal{E}_4|\mathcal{F}) &= P(\exists (i, j, k) \neq (G_C[l], G_A[l], G_U[l-1]) \text{ s.t. } \psi(i, j, k) = 1 | \mathcal{F}) \\ &= \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4,\end{aligned}\quad (84)$$

where

$$\Phi_1 = \Pr(\exists j \neq G_A[l] \text{ s.t. } \psi(G_C[l], j, G_U[l-1]) = 1 | \mathcal{F}), \quad (85)$$

$$\Phi_2 = \Pr(\exists i \neq G_C[l], j \text{ s.t. } \psi(i, j, G_U[l-1]) = 1 | \mathcal{F}), \quad (86)$$

$$\Phi_3 = \Pr(\exists k \neq G_U[l-1], j \text{ s.t. } \psi(G_C[l], j, k) = 1 | \mathcal{F}), \quad (87)$$

$$\Phi_4 = \Pr(\exists i \neq G_C[l], k \neq G_U[l-1], j \text{ s.t. } \psi(i, j, k) = 1 | \mathcal{F}). \quad (88)$$

### C.1 Upper bound for $\Phi_1$

Using the union bound, we have

$$\begin{aligned}\Phi_1 &\leq \sum_{j=1}^{2^{n\rho_1}} \Pr [ (\mathbf{U}[l-1], \mathbf{A}[l-1], \mathbf{Y}[l-1], \mathbf{C}[l-1], \mathbf{C}[l], \mathbf{Y}[l], \mathbf{A}_{[l, j, \mathbf{U}[l-1], \mathbf{C}[l-1], \mathbf{C}[l]]}) \\ &\quad \in \mathcal{A}_{\epsilon_l}^{(n)}(P_{\tilde{C}\tilde{A}\tilde{U}\tilde{Y}CA\tilde{Y}}), G_A[l] \neq j | \mathcal{F} ].\end{aligned}\quad (89)$$

For brevity, we denote the tuple  $(\mathbf{U}[l-1], \mathbf{C}[l-1], \mathbf{C}[l])$  by  $\mathbf{T}'$  and the tuple  $(\mathbf{U}[l-1], \mathbf{C}[l-1], \mathbf{C}[l], \mathbf{A}[l-1], \mathbf{Y}[l-1])$  by  $\mathbf{T}$ . (89) can then be written as

$$\begin{aligned}\Phi_1 &\leq \frac{1}{\Pr[\mathcal{F}]} \sum_{j=1}^{2^{n\rho_1}} \sum_{\mathbf{t}, \mathbf{a}, \mathbf{y} \in \mathcal{A}_{\epsilon_l}} \Pr[\mathbf{T} = \mathbf{t}, \mathbf{A}_{[l,j,\mathbf{T}']} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y}, G_A[l] \neq j] \\ &= \frac{2^{n\rho_1}}{\Pr[\mathcal{F}]} \sum_{\mathbf{t}, \mathbf{a}, \mathbf{y} \in \mathcal{A}_{\epsilon_l}} \Pr[\mathbf{T} = \mathbf{t}] \Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y}, G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}]\end{aligned}\tag{90}$$

where the second equality is due to the symmetry of the codebook construction. We note that the index  $G_A[l]$  is a *function* of the entire  $A$ -codebook  $\{\mathbf{A}_{[l,j,\mathbf{U}[l-1],\mathbf{C}[l-1],\mathbf{C}[l]]}, j = 1 \dots 2^{n\rho_1}\}$  and so the events

$$G_A[l] \neq 1 \text{ and } (\mathbf{A}_{[l,1,\mathbf{U}[l-1],\mathbf{C}[l-1],\mathbf{C}[l]]} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y})$$

are dependent. This dependency can be handled using the technique developed in [28].

Let  $\bar{\mathcal{C}}$  be the set  $\{\mathbf{A}_{[l,j,\mathbf{U}[l-1],\mathbf{C}[l-1],\mathbf{C}[l]]}, j = 2 \dots 2^{n\rho_1}\}$ , i.e.,  $\bar{\mathcal{C}}$  is the  $A$ -codebook without the first codeword. Focusing on the inner term of the summation in (90), we have

$$\begin{aligned}&\Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y}, G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}] \\ &\leq \Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &= \sum_{\bar{\mathbf{c}}} \Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{\mathbf{c}}] \cdot \Pr[\bar{\mathcal{C}} = \bar{\mathbf{c}} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &\stackrel{(a)}{=} \sum_{\bar{\mathbf{c}}} \Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{\mathbf{c}}] \cdot \Pr[\mathbf{Y}[l] = \mathbf{y} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{\mathbf{c}}] \cdot \Pr[\bar{\mathcal{C}} = \bar{\mathbf{c}} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &\stackrel{(b)}{\leq} 2 \cdot \Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid \mathbf{T}' = \mathbf{t}'] \sum_{\bar{\mathbf{c}}} \Pr[\mathbf{Y}[l] = \mathbf{y} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{\mathbf{c}}] \Pr[\bar{\mathcal{C}} = \bar{\mathbf{c}} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &\stackrel{(c)}{\leq} 4 \cdot \Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid \mathbf{T}' = \mathbf{t}'] \Pr[\mathbf{Y}[l] = \mathbf{y} \mid \mathbf{T} = \mathbf{t}] \\ &\stackrel{(d)}{\leq} 4 \cdot 2^{-n(H(A|C\bar{C}\bar{U}) - \delta(\epsilon_l))} \cdot 2^{-n(H(Y|C\bar{C}\bar{U}\bar{A}\bar{Y}) - \delta(\epsilon_l))}.\end{aligned}\tag{91}$$

In the chain above, (a) is true because given  $G_A[l] \neq 1$ , the following Markov chain holds:

$$\mathbf{A}_{[l,1,\mathbf{T}']} - (\bar{\mathcal{C}}, \mathbf{T}) - \mathbf{X}[l] - \mathbf{Y}[l].$$

(d) follows from Property 3 of typical sequences, while (b) and (c) are obtained from the following claim, proved along the lines of [28, Lemmas 1 and 2].

**Claim 1.** 1.  $\Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{\mathbf{c}}] \leq 2 \Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid \mathbf{T}' = \mathbf{t}']$ .  
2.  $\Pr[\mathbf{Y}[l] = \mathbf{y} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}] \leq 2 \Pr[\mathbf{Y}[l] = \mathbf{y} \mid \mathbf{T} = \mathbf{t}]$ .

*Proof.* We have

$$\begin{aligned}
\Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}] &= \frac{\Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a}, G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}]}{\Pr[G_A[l] \neq 1, \mid \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}]} \\
&\leq \frac{\Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}]}{\Pr[G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}]} \\
&= \frac{\Pr[\mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid \mathbf{T}' = \mathbf{t}']}{\Pr[G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}]}
\end{aligned} \tag{92}$$

where the last equality holds because each codeword of the codebook  $\{\mathbf{A}_{[l,j,\mathbf{T}']}, j = 1, \dots, 2^{n\rho_1}\}$  is independently generated, conditioned only on the symbols of  $\mathbf{T}'$ . We now provide a lower bound for the denominator of (92).

$$\begin{aligned}
\Pr[G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}] &= 1 - \Pr[G_A[l] = 1 \mid \mathbf{T} = \mathbf{t}, \bar{\mathcal{C}} = \bar{c}] \\
&\geq 1 - \Pr\left[(\mathbf{A}_{[l,1,\mathbf{T}']}, \mathbf{T}) \in \mathcal{A}_{\epsilon_l}^{(n)}(P_{\tilde{C}\tilde{A}\tilde{U}\tilde{Y}CA})\right] \\
&\geq 1 - 2^{-n(I(A;\tilde{A}\tilde{Y}|\tilde{U}\tilde{C}C) - \delta(\epsilon_l))} \\
&\geq \frac{1}{2}
\end{aligned} \tag{93}$$

for sufficiently large  $n$ . Substituting in (92) completes the proof of the first part of the claim.

For the second part, we write

$$\begin{aligned}
\Pr[\mathbf{Y}[l] = \mathbf{y} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}] &= \frac{\Pr[\mathbf{Y}[l] = \mathbf{y}, G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}]}{\Pr[G_A[l] \neq 1, \mid \mathbf{T} = \mathbf{t}]} \\
&\leq \frac{\Pr[\mathbf{Y}[l] = \mathbf{y} \mid \mathbf{T} = \mathbf{t}]}{\Pr[G_A[l] \neq 1 \mid \mathbf{T} = \mathbf{t}]} \\
&= \frac{\Pr[\mathbf{Y}[l] = \mathbf{y} \mid \mathbf{T} = \mathbf{t}]}{(2^{n\rho_1} - 1)/2^{n\rho_1}} \\
&\leq 2 \cdot \Pr[\mathbf{Y}[l] = \mathbf{y} \mid \mathbf{T} = \mathbf{t}]
\end{aligned} \tag{94}$$

for large enough  $n$ . The second equality in the chain above due to the symmetry of the codebook construction. This completes the proof of the claim.  $\square$

Substituting the bound from (91) in (90), we obtain

$$\begin{aligned}
\Phi_1 &\leq \frac{2^{n\rho_1}}{\Pr[\mathcal{F}]} \sum_{\mathbf{t}} \Pr[\mathbf{T} = \mathbf{t}] \sum_{\mathbf{a}, \mathbf{y} \in \mathcal{A}_{\epsilon_l}(\cdot, \mathbf{t})} 4 \cdot 2^{2n\delta(\epsilon_l)} \cdot 2^{-nH(A|C\tilde{C}\tilde{U})} \cdot 2^{-nH(Y|C\tilde{C}\tilde{U}\tilde{A}\tilde{Y})} \\
&\stackrel{(a)}{\leq} \frac{2^{n\rho_1}}{\Pr[\mathcal{F}]} \sum_{\mathbf{t}} \Pr[\mathbf{T} = \mathbf{t}] 2^{n(H(AY|C\tilde{C}\tilde{U}\tilde{A}\tilde{Y}) + \delta(\epsilon_l))} \left(4 \cdot 2^{2n\delta(\epsilon_l)} \cdot 2^{-nH(A|C\tilde{C}\tilde{U})} \cdot 2^{-nH(Y|C\tilde{C}\tilde{U}\tilde{A}\tilde{Y})}\right) \\
&= \frac{4 \cdot 2^{3n\delta(\epsilon_l)} \cdot 2^{n\rho_1} \cdot 2^{nH(AY|C\tilde{C}\tilde{U}\tilde{A}\tilde{Y})}}{\Pr[\mathcal{F}] \cdot 2^{nH(A|C\tilde{C}\tilde{U})} \cdot 2^{nH(Y|C\tilde{C}\tilde{U}\tilde{A}\tilde{Y})}}
\end{aligned} \tag{95}$$

where (a) follows from the upper bound on the size of the conditionally typical set (Property 2).

## C.2 Upper bound for $\Phi_2, \Phi_3, \Phi_4$

Using the union bound, we have

$$\begin{aligned} \Phi_2 \leq \sum_{i=1}^{2^{n\rho_0}} \sum_{j=1}^{2^{n\rho_1}} \Pr \Big[ (\mathbf{U}[l-1], \mathbf{A}[l-1], \mathbf{Y}[l-1], \mathbf{C}[l-1], \mathbf{Y}[l], \mathbf{C}_{[l,i,\mathbf{C}[l-1]]}, \mathbf{A}_{[l,j,\mathbf{U}[l-1],\mathbf{C}[l-1],\mathbf{C}_{[l,i,\mathbf{C}[l-1]]}]}]) \\ \in \mathcal{A}_{\epsilon_l}^{(n)}(P_{\tilde{C}\tilde{A}\tilde{U}\tilde{Y}CAY}), G_C[l] \neq i \mid \mathcal{F} \Big]. \end{aligned} \quad (96)$$

To keep the notation manageable, in the next few equations we will use the shorthand  $\mathbf{C}_i$  for  $\mathbf{C}_{[l,i,\mathbf{C}[l-1]]}$ . We also redefine  $\mathbf{T}'$  as the tuple  $(\mathbf{U}[l-1], \mathbf{C}[l-1])$  and  $\mathbf{T}$  as the tuple  $(\mathbf{U}[l-1], \mathbf{C}[l-1], \mathbf{A}[l-1], \mathbf{Y}[l-1])$ . (96) can then be written as

$$\begin{aligned} \Phi_2 &\leq \frac{1}{\Pr[\mathcal{F}]} \sum_{i=1}^{2^{n\rho_0}} \sum_{j=1}^{2^{n\rho_1}} \sum_{\mathbf{t}, \mathbf{c}, \mathbf{a}, \mathbf{y} \in \mathcal{A}_{\epsilon_l}} \Pr[\mathbf{T} = \mathbf{t}, \mathbf{C}_i = \mathbf{c}, \mathbf{A}_{[l,j,\mathbf{T}',\mathbf{C}_i]} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y}, G_C[l] \neq i] \\ &= \frac{2^{n(\rho_0+\rho_1)}}{\Pr[\mathcal{F}]} \sum_{\mathbf{t}, \mathbf{a}, \mathbf{c}, \mathbf{y} \in \mathcal{A}_{\epsilon_l}} \Pr[\mathbf{T} = \mathbf{t}] \Pr[\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{T}',\mathbf{C}_1]} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y}, G_C[l] \neq 1 \mid \mathbf{T} = \mathbf{t}] \end{aligned} \quad (97)$$

where the second equality is due to the symmetry of the codebook construction. We note that the index  $G_C[l]$  is a function of the entire  $C$ -codebook  $\{\mathbf{C}_i = \mathbf{C}_{[l,i,\mathbf{C}[l-1]]}, i = 1 \dots 2^{n\rho_0}\}$  and so the events

$$G_C[l] \neq 1 \text{ and } (\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{U}[l-1],\mathbf{C}[l-1],\mathbf{C}_1]} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y})$$

are dependent. Define  $\bar{\mathcal{C}}$  as  $\{\mathbf{C}_i = \mathbf{C}_{[l,i,\mathbf{C}[l-1]]}, i = 2 \dots 2^{n\rho_0}\}$ , i.e., the  $C$ -codebook without the first codeword. We then have

$$\begin{aligned} &\Pr[\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{T}',\mathbf{C}_1]} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y}, G_C[l] \neq 1 \mid \mathbf{T} = \mathbf{t}] \\ &\leq \Pr[\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{T}',\mathbf{C}_1]} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y} \mid G_C[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &= \sum_{\bar{\mathbf{c}}} \Pr[\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a}, \mathbf{Y}[l] = \mathbf{y} \mid G_C[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathbf{C}} = \bar{\mathbf{c}}] \cdot \Pr[\bar{\mathbf{C}} = \bar{\mathbf{c}} \mid G_C[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &\stackrel{(a)}{=} \sum_{\bar{\mathbf{c}}} \Pr[\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid G_C[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathbf{C}} = \bar{\mathbf{c}}] \cdot \Pr[\mathbf{Y}[l] = \mathbf{y} \mid G_C[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathbf{C}} = \bar{\mathbf{c}}] \cdot \Pr[\bar{\mathbf{C}} = \bar{\mathbf{c}} \mid G_C[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &\stackrel{(b)}{\leq} 2 \cdot \Pr[\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid \mathbf{T}' = \mathbf{t}'] \sum_{\bar{\mathbf{c}}} \Pr[\mathbf{Y}[l] = \mathbf{y} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}, \bar{\mathbf{C}} = \bar{\mathbf{c}}] \Pr[\bar{\mathbf{C}} = \bar{\mathbf{c}} \mid G_A[l] \neq 1, \mathbf{T} = \mathbf{t}] \\ &\stackrel{(c)}{\leq} 4 \cdot \Pr[\mathbf{C}_1 = \mathbf{c}, \mathbf{A}_{[l,1,\mathbf{T}']} = \mathbf{a} \mid \mathbf{T}' = \mathbf{t}'] \Pr[\mathbf{Y}[l] = \mathbf{y} \mid \mathbf{T} = \mathbf{t}] \\ &\stackrel{(d)}{\leq} 4 \cdot 2^{-n(H(C|\tilde{C})+H(A|C\tilde{C}\tilde{U})-\delta(\epsilon_l))} \cdot 2^{-n(H(Y|\tilde{C}\tilde{U}\tilde{A}\tilde{Y})-\delta(\epsilon_l))}. \end{aligned} \quad (98)$$

In the chain above, (a) is true because given  $G_C[l] \neq 1$ , the following Markov chain holds:

$$(\mathbf{C}_{[l,1,\mathbf{C}[l-1]]}, \mathbf{A}_{[l,1,\mathbf{T}']}) - (\bar{\mathbf{C}}, \mathbf{T}) - \mathbf{X}[l] - \mathbf{Y}[l].$$

(d) follows from Property 3 of typical sequences, while (b) and (c) follow from arguments very similar to Claim



1 in the previous subsection. Substituting the bound from (98) in (97), we obtain

$$\begin{aligned}
\Phi_2 &\leq \frac{2^{n(\rho_0+\rho_1)}}{\Pr[\mathcal{F}]} \sum_{\mathbf{t}} \Pr[\mathbf{T} = \mathbf{t}] \sum_{\mathbf{c}, \mathbf{a}, \mathbf{y} \in \mathcal{A}_{\epsilon_l}(\cdot|\mathbf{t})} 4 \cdot 2^{2n\delta(\epsilon_l)} \cdot 2^{-nH(C|\tilde{C})} \cdot 2^{-nH(A|C\tilde{C}\tilde{U})} \cdot 2^{-nH(Y|\tilde{C}\tilde{U}\tilde{A}\tilde{Y})} \\
&\stackrel{(a)}{\leq} \frac{2^{n(\rho_0+\rho_1)}}{\Pr[\mathcal{F}]} \sum_{\mathbf{t}} \Pr[\mathbf{T} = \mathbf{t}] 2^{n(H(AYC|\tilde{C}\tilde{U}\tilde{A}\tilde{Y})+\delta(\epsilon_l))} \left( 4 \cdot 2^{2n\delta(\epsilon_l)} \cdot 2^{-nH(C|\tilde{C})} \cdot 2^{-nH(A|C\tilde{C}\tilde{U})} \cdot 2^{-nH(Y|\tilde{C}\tilde{U}\tilde{A}\tilde{Y})} \right) \\
&= \frac{4 \cdot 2^{3n\delta(\epsilon_l)} \cdot 2^{n(\rho_1+\rho_0)} \cdot 2^{nH(AYC|\tilde{C}\tilde{U}\tilde{A}\tilde{Y})}}{\Pr[\mathcal{F}] \cdot 2^{nH(C|\tilde{C})} \cdot 2^{nH(A|C\tilde{C}\tilde{U})} \cdot 2^{nH(Y|\tilde{C}\tilde{U}\tilde{A}\tilde{Y})}}
\end{aligned} \tag{99}$$

where (a) follows from the upper bound on the size of the conditionally typical set.

In a similar fashion, we can obtain the following bounds for  $\Phi_3$  and  $\Phi_4$ .

$$\Phi_3 \leq \frac{4 \cdot 2^{3n\delta(\epsilon_l)} \cdot 2^{n(R'_1+\rho_1)} \cdot 2^{nH(\tilde{U}AY|\tilde{C}\tilde{A}\tilde{Y})}}{\Pr[\mathcal{F}] \cdot 2^{nH(\tilde{U}|\tilde{C})} \cdot 2^{nH(A|C\tilde{C}\tilde{U})} \cdot 2^{nH(Y|C\tilde{C}\tilde{A}\tilde{Y})}}, \tag{100}$$

$$\Phi_4 \leq \frac{4 \cdot 2^{3n\delta(\epsilon_l)} \cdot 2^{n(R'_1+\rho_0+\rho_1)} \cdot 2^{nH(\tilde{U}CAY|\tilde{C}\tilde{A}\tilde{Y})}}{\Pr[\mathcal{F}] \cdot 2^{nH(\tilde{U}|\tilde{C})} \cdot 2^{nH(C|\tilde{C})} \cdot 2^{nH(A|C\tilde{C}\tilde{U})} \cdot 2^{nH(Y|\tilde{C}\tilde{A}\tilde{Y})}}. \tag{101}$$

Lemmas 5.1 and 5.2 together with the induction hypothesis that  $\Pr[\mathcal{E}_4[k]] < \epsilon$  for  $k = 1, \dots, l-1$  imply that  $\Pr[\mathcal{F}] > 1 - 5\epsilon l$ , which is close to 1 for  $\epsilon \ll 1/L$ . Thus the bounds (95), (99), (100) and (101) can be made arbitrarily small for sufficiently large  $n$  if the conditions of the lemma are satisfied.

Substituting back in (84), we obtain  $P[\mathcal{E}_4[l] \mid \mathcal{F}] \leq \epsilon$  for all sufficiently large  $n$ . Similarly, one can show that  $P[\mathcal{E}_5[l] \mid \mathcal{F}] \leq \epsilon$  if the conditions in the statement of lemma are satisfied.